EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), 2014. Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 2 (surveillance and data management activities)

EFSA Publication

Link to article, DOI:
10.2903/j.efsa.2014.3784

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

European Food Safety Authority

EFSA Journal 2014;12(7):3784

SCIENTIFIC OPINION

Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 2 (surveillance and data management activities)[1]

EFSA Panel on Biological Hazards (BIOHAZ)[2,3]

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

Surveillance programmes based on active and harmonised sampling are considered the most suitable for food-borne outbreak investigations, hypothesis generation, early detection of emerging pathogen subtypes, attribution modelling and genetic studies of bacterial populations. Currently, prototype molecular databases are not widely linked and contain limited epidemiological data, therefore development of linkage mechanisms is a priority. A key technical requirement is determination of an agreed threshold value for the level of genetic variation amongst isolates that can still be regarded as epidemiologically-related. Molecular typing data should be coupled with a minimum required set of epidemiological data and datasets should be comparable to facilitate joint analyses in conjunction with human case data. Rules for assembling strain collections and associated provenance data should be agreed and introduced as EU standards. The data collection process and the characteristics of the data repository should ensure reproducibility and maximise compatibility and interoperability between different datasets. Molecular bacterial characterisation developments, particularly Whole Genome Sequencing (WGS), should be harmonised with those used for surveillance in the human population and food industry. Reference methods and materials, including sequence data, should be adopted for typing of food-borne pathogens. Upload of molecular data should only be allowed for approved laboratories and should be subject to External Quality Assessment. Ongoing international oversight is required to ensure a consensual 'one-health' approach. The establishment of a joint EFSA-ECDC-EU-RLs committee for the support of cross-sectoral molecular surveillance, with a balance of public health and veterinary expertise and including both epidemiologists and microbiologists is strongly recommended. Revision of the legal basis of programmes for pathogen reduction based on historic organism nomenclature may be necessary following the increased use of WGS and the subsequent identification of more biologically relevant groupings of organisms.

KEY WORDS

genotyping, molecular typing, whole genome sequencing, surveillance, data management

---

[1] On request from EFSA, Question No EFSA-Q-2013-00906, adopted on 10 July 2014.

[2] Panel members: Olivier Andreoletti, Dorte Lau Baggesen, Declan Bolton, Patrick Butaye, Paul Cook, Robert Davies, Pablo S. Fernandez Escamez, John Griffin, Tine Hald, Arie Havelaar, Kostas Koutsoumanis, Roland Lindqvist, James McLauchlin, Truls Nesbakken, Miguel Prieto Maradona, Antonia Ricci, Giuseppe Ru, Moez Sanaa, Marion Simmons, John Sofos and John Threlfall. Correspondence: biohaz@efsa.europa.eu

[3] Acknowledgement: The Panel wishes to thank the members of the Working Group on the evaluation of molecular typing methods for major food-borne pathogens: Dorte Lau Baggesen, Patrick Butaye, Robert Davies, Tine Hald, Arie Havelaar, Bjørn-Arne Lindstedt, Martin Maiden, Eva Møller Nielsen, Gaia Scavia and John Threlfall for the preparatory work on this scientific opinion and European Centre for Disease Prevention and Control (ECDC) staff: Marc Struelens, and EFSA staff: Maria Teresa da Silva Felício, Ernesto Liebana Criado and Luis Vivas-Alegre for the support provided to this scientific opinion.

Suggested citation: EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), 2014. Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 2 (surveillance and data management activities). EFSA Journal 2014;12(7):3784, 46 pp. doi:10.2903/j.efsa.2014.3784

Available online: www.efsa.europa.eu/efsajournal

SUMMARY

The European Food Safety Authority (EFSA) asked the Panel on Biological Hazards (BIOHAZ) to deliver a scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards (i.e. *Salmonella*, thermophilic *Campylobacter*, Shiga toxin-producing *Escherichia coli* (STEC) and *Listeria monocytogenes*) and their use for attribution modelling, outbreak investigation and surveillance, including data-related issues.

Following that request, the BIOHAZ Panel adopted, on 5 December 2013, a Scientific Opinion addressing the evaluation of molecular typing methods and their suitability for the different applications that were considered (EFSA, 2013a). Important conclusions of that Opinion were that data from strain characterisation should be linked with epidemiological data, that the selection of isolates must be unbiased and statistically representative of the population to be assessed and that international harmonisation of molecular characterisation outputs by means of standardisation or appropriate quality control procedures is essential. Important recommendations were that cross-sector (humans, food, food animals and related environments) and international coordination of method validation is required as a priority, and that development and improvement of international initiatives with regard to harmonised platforms for sharing of data should be urgently prioritized, including the integration of Whole Genome Sequencing (WGS) into such international platforms.

In the current scientific Opinion, the BIOHAZ Panel has addressed data-related issues, in particular: (i) the evaluation of the requirements for the design of surveillance activities for food-borne pathogens, especially regarding the selection of statistically representative group of isolates to be included in molecular typing investigations and attribution modelling; and (ii) the requirements for harmonised data collection, management and analysis, with the final aim of achieving full integration of efficient and effectively managed molecular typing databases for food-borne pathogens. In order to provide a comprehensive overview of the applicability of molecular typing methods for the aforementioned food-borne pathogens in the given applications, both the Opinions should be referred to.

In the scope of this Opinion, the term 'monitoring' has been applied to describe a system of collecting, analyzing and disseminating data on the occurrence of zoonoses, zoonotic agents and antimicrobial resistance of public health relevance in the food chain. 'Surveillance' is understood as the systematic ongoing collection, collation and analysis of information related to food safety and the timely dissemination of information to appropriate persons so that action can be taken. Public health surveillance has been defined as the ongoing, systematic collection, analysis and interpretation of health data, essential to the planning, implementation and evaluation of public health practice, closely integrated with the dissemination of these data to appropriate persons and linked to prevention and control.

Surveillance programmes based on active and harmonised sampling are most suitable for statistical analysis which may be used for testing hypotheses. They provide the most complete, accurate and representative data and are more likely to be suitable for source attribution and detailed/advanced epidemiological investigations and risk assessments, as long as the datasets are sufficiently large to support robust statistical analyses. Typing results of isolates collected from routine laboratory submissions where the isolates are linked to limited information can still be valuable and may help support food-borne outbreak investigations, generation of hypotheses, early detection of emerging pathogen subtypes and genetic studies of bacterial populations, but sampling bias should be taken into account when formulating conclusions.

The introduction of molecular typing-based surveillance should include the establishment of a continuous information cycle to provide accurate and representative data over time and space, to include the relevant typing characteristics of specified food-borne pathogens (i.e. *Salmonella*, STEC, *L. monocytogenes* and thermophilic *Campylobacter* spp.) in food animal species and key points in the

food production chain. Currently, various non-comparable methods are applied for the molecular typing of these pathogens worldwide. Pulsed-field gel electrophoresis (PFGE), is still the most widely used method for subtyping of *Salmonella*, STEC and *L. monocytogenes*. For *S.* Typhimurium and *S.* Enteritidis, PFGE may be used together with Multi-Locus Variable number tandem repeat Analysis (MLVA); although MLVA is increasingly being used as the sole method. Multi-locus sequence typing (MLST) has been the method of choice for thermophilic *Campylobacter* but is being superseded by WGS. Routine molecular typing of *Campylobacter jejuni/coli* has not been shown to add value for outbreak detection but may contribute to source attribution studies for campylobacteriosis.

Integrated analyses will be optimised if surveillance activities incorporate complete datasets containing all relevant information on the isolate. Examples of such datasets are those related to the genotype and other characteristics such as serovar or antimicrobial resistance profile, coupled with accurate data on the effect on the host and related epidemiological data. At present, prototype databases cannot be used for surveillance purposes since they are not widely linked to epidemiological data. Thus, the development of linkage mechanisms to access complex genetic and epidemiological data within different databases may be required.

A key priority in relation to integrated public health surveillance is to determine a threshold value for the level of genetic variation amongst isolates that can still be regarded as epidemiologically related. This threshold will vary according to the organism under investigation, time frame, population size and geographical scope of the investigation of the chain of transmission. The discriminatory power of a method describes its capacity to assign different subtypes to epidemiologically unrelated strains in the population studied, and is thereby a tool for describing the threshold for separation of epidemiologically related and unrelated isolates. A high discriminatory power will often lead to the division of panels of isolates into many subtypes, where the probability of categorizing unrelated isolates to the same subtype is small. With increasing discriminatory power, the probability of assigning related strains to different subtypes may also increase. In contrast, a relatively low discriminatory power will result in fewer subtypes and the probability of categorizing related isolates to different subtypes is small, but the probability of including unrelated isolates in the defined subtype is likely to increase. In the integrated analysis of typing data and epidemiological data it is important to optimise the discriminatory power/threshold for separation in a way which gives the most meaningful grouping of isolates from an epidemiological perspective to obtain the highest level of epidemiological concordance.

The collection of data for molecular typing-based surveillance of food-borne pathogens in animals, feed and food should be based on active sampling and an agreed sampling design should be prioritized for the purposes of molecular surveillance of pathogens in the food chain and from human cases. The use of alternative sources of data and strains should be carefully evaluated according to the required outcome and to a set of established criteria. The applied molecular typing methods should be based on both the pathogen to be characterised and the level of discriminatory power required depending on the required application of the surveillance results. Furthermore, molecular typing data should be coupled with a minimum required set of epidemiological data including, for example, information on the sampling context and population/sample set under study. Datasets generated should be comparable and suitable for joint analysis with other data from parallel surveillance in humans and/or relevant samples. Surveillance activities should be primarily aimed at investigating the priority source/pathogen combinations and be robust and statistically based. Rules for assembling strain collections and associated provenance data from general surveillance of pathogens should be agreed and introduced as EU standards.

When assessing requirements for integrated and harmonised data collection and management activities, the data collection process and the characteristics of the data repository should ensure the highest level of both the reproducibility of data and analyses, over time and space, and maximise the compatibility and interoperability among different datasets. This would be best accomplished by providing the overall architecture of a surveillance programme that includes the highest level of harmonisation with either international standards, if available, or a uniform approach to collection,

management and analysis of data. Opportunities for harmonisation are facilitated by European Union Reference Laboratories (EU-RLs) which have an important role to support harmonisation in the laboratory characterization of food-borne hazards and active involvement in coordination of development and implementation of new molecular typing methods will be an important priority within the remit of EU-RLs in future years. Development of molecular methods for characterisation of food-borne pathogenic bacteria in animals, feed and food should be harmonised with those adopted for the surveillance of similar food-borne pathogens in the human population. Reference methods and materials, including sequence data, should be adopted for typing characterization of food-borne pathogens, and upload of data should be allowed only for approved laboratories.

Since the rapid development of sequence-based methodology is likely to outstrip the capabilities of individual centres of expertise, ongoing international expert consultation and oversight is required to optimise the opportunities offered by WGS. This should involve specialist centres, specialist scientists, bioinformaticians, risk assessors and risk managers from public health, veterinary, food production and retail sectors to identify issues and design a consensual 'one health' approach. Finally, the BIOHAZ Panel strongly recommends the establishment of a joint EFSA-ECDC-EU-RLs committee for the support of cross-sectoral molecular surveillance, to represent a balance of expertise from the public health and veterinary/food sectors as well as epidemiologists and microbiologists.

TABLE OF CONTENTS

BACKGROUND AS PROVIDED BY EFSA

It is important to link closely molecular surveillance initiatives instigated for pathogens identified in the human population and surveillance activities in food, feed and food-producing animals. This would help to identify common sources of infection for the animals themselves, e.g. via internationally-traded feed ingredients and replacement breeding and commercial stock, and would provide a means of comparing human and animal strains via real time surveillance and as part of outbreak investigations.

A wide variety of sub-typing methods exist for most pathogens but they are often applied in a way that is not standardised and dependent on individual protocols, approaches and equipment used in separate laboratories. The introduction of harmonised protocols and reference strains e.g. for pulsed field gel electrophoresis (PFGE), and for Multiple-Locus Variable Number Tandem Repeat Analysis (MLVA) as part of the PulseNet[4] initiative represent an attempt to introduce harmonisation of methodology or standardisation of interpretation. PulseNet in particular has been particularly valuable in the USA, identifying numerous diffuse common source outbreaks of *Salmonella* spp. or STEC[5] that would otherwise have been considered to be sporadic cases. The identification of such outbreaks allows interventions such as product recall that can shorten the duration of food-borne disease outbreaks and potentially save lives. Furthermore, by identifying the factors that caused the outbreak, HACCP plans and food safety standards may be reviewed, helping to reduce future outbreaks or sporadic cases.

In recent years EFSA has made increasing use of attribution modelling to enhance the scientific value of Opinions. This approach has been very valuable to help risk managers focus regulatory attention on the highest priority sources of food-borne infection. The precision of attribution modelling based on sub-typing of organisms is limited both by the scarcity of harmonised data for some food animal species, e.g. for *Salmonella* spp. in the bovine reservoir, and the occurrence of similar organisms at the serovar level in different animal populations. In the case of other organisms such as thermophilic *Campylobacter*, even this level of sub-typing detail is largely lacking. Various studies have shown that in many cases further distinction between sources, both in terms of animal reservoir and geographical origin can be made by inclusion of additional combinations of phenotypic or molecular sub-typing data. A notable example of this is the use of multi-locus sequence typing (MLST) for thermophilic *Campylobacter* in studies in New Zealand and UK. It has recently been demonstrated that the use of MLST typing data in combination with case-control studies can provide novel perspectives on the risk factors for human disease in relation to different animal reservoirs.

DNA sequence-based approaches, including whole genome sequencing (WGS), have prepared the stage for future revolutionary advances in diagnostic and typing techniques. Increasing use of data generated from next-generation sequencing (NGS) technologies is expected to provide the means for a paradigm shift in the way microorganisms are identified and characterised. This will result in a much greater ability to undertake detailed analysis and more rapidly identify dispersed outbreaks, such as those arising from national or international distribution of contaminated foods. Epigenetic techniques and quantitative gene expression arrays may also in the future be used to provide early indication of potential new and emerging epidemic strains.

Harmonised approaches for (i) selection of representative isolates of food-borne pathogens, (ii) selection of sub-typing methodologies, and (iii) analysis and storage of large quantities of molecular typing data, would facilitate provision of valuable guidance from EFSA to the scientific community and regulatory bodies, particularly in the areas of outbreak detection and source attribution modelling for food-borne pathogens. To that end it is the intention to include participation of ECDC and EU Reference laboratories in this working group. Such an approach would enhance the value and integration of current molecular typing schemes and should ultimately assist in the application of improved tools to further enhance the protection of public health.

[4] Further information on PulseNet International available at: http://www.pulsenetinternational.org/ (last visited on 11/12/2013)

[5] Shiga toxin-producing *E. coli*

TERMS OF REFERENCE AS PROVIDED BY EFSA

EFSA requests the BIOHAZ Panel to:

1. Review information on current and prospective (e.g. WGS) molecular identification and sub-typing methods for food-borne pathogens (e.g. *Salmonella*, thermophilic *Campylobacter*, STEC and *Listeria*) in terms of discriminatory capability, reproducibility, and capability for international harmonisation.

2. Review the appropriateness of use of the different food-borne pathogen sub-typing methodologies (including data analysis methods) for outbreak investigation, attribution modelling and the potential for early identification of organisms with future epidemic potential.

3. **Evaluate the requirements for the design of surveillance activities for food-borne pathogens, in particular for the selection for a statistically representative group of isolates to be included in molecular typing investigations, and attribution modelling.**

4. **Review the requirements for harmonised data collection, management and analysis, with the final aim to achieve full integration of efficient and effectively managed molecular typing databases for food-borne pathogens.**

Following a proposal made by the BIOHAZ Panel, EFSA agreed upon the delivery of two separate Scientific Opinions: one covering Terms of Reference one and two (adopted by the Panel on 5 December 2013[6]), and the Opinion presented here, covering Terms of Reference three and four.

[6] EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), 2013. Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 1 (evaluation of methods and applications). EFSA Journal 2013;11(12):3502, 84 pp. doi:10.2903/j.efsa.2013.3502

ASSESSMENT

1. Introduction

The Panel on Biological Hazards (BIOHAZ) adopted on 5 December 2013 a scientific Opinion addressing the evaluation of molecular typing methods and their suitability for the different applications sought. In the current scientific Opinion the BIOHAZ Panel addresses data-related issues and in particular: (i) the evaluation of the requirements for the design of surveillance activities for food-borne pathogens, especially for the selection of a statistically representative group of isolates to be included in molecular typing investigations and attribution modelling; and (ii) the review of the requirements for harmonised data collection, management and analysis, with the final aim of achieving full integration of efficient and effectively managed molecular typing databases for food-borne bacterial pathogens. In order to provide a comprehensive overview of the general applicability of molecular typing methods for *Salmonella*, thermophilic *Campylobacter*, Shiga toxin-producing *Escherichia coli* (STEC) and *Listeria monocytogenes*, both of these Opinions should be consulted.

According to the European Centre for Disease Prevention and Control (ECDC), molecular typing refers to the application of laboratory methods capable of characterising, discriminating and indexing subtypes of microorganisms and thereby supporting epidemiological studies of their source, distribution and spread (ECDC, 2007, 2013b; EFSA, 2013a). The typing is, as a rule, applied for characterisation below species level. The nomenclature employed is less well defined and varies between different genera (van Belkum et al., 2007) in contrast to taxonomic classification, which is governed by the International Code of Nomenclature of Bacteria (Lapage et al., 1992).

In the first part of the Opinion (EFSA, 2013a) an important conclusion was that "*data from strain characterisation should be linked with epidemiological data and the strain selection must be unbiased and statistically representative of the population to be assessed*". A further conclusion was that: "*international harmonisation of molecular characterisation outputs by means of standardisation or appropriate quality control procedures is essential*". An important recommendation was that "*cross-sector (humans, food, food animals and related environments) and international coordination of method validation is required as a priority*". Furthermore, "development and improvement of international initiatives with regard to harmonized platforms for sharing of data such as those promoted by PulseNet and ECDC/EFSA should be urgently prioritized, including the integration of whole genome sequencing (WGS) into such international platforms." The outcome of the first part of the Opinion therefore addresses the need for coordination and collaboration across sectors in relation to the establishment and development of the integrated surveillance in humans, animals, feed and food.

Both 'surveillance' and 'monitoring' as they are defined in the veterinary field (Noordhuizen et al., 2001; Salman et al., 2003) or by international bodies, such as the World Organisation for Animal Health (OIE) (Hassan, 2007) rely on "*the ongoing and/or repetitive process of sampling individuals from an animal population and food/feed sources to assess their health status or a particular event over time and space*".

In the scope of this Opinion, and in agreement with Directive 2003/99/EC[7], the term 'monitoring' will be applied to describe a system of collecting, analyzing and disseminating data on the occurrence of zoonoses, zoonotic agents and antimicrobial resistance related thereto. 'Surveillance' is understood as the systematic ongoing collection, collation and analysis of information related to food safety and the timely dissemination of information to appropriate persons so that action can be taken. Public health surveillance has been defined as the ongoing, systematic collection, analysis and interpretation of health data essential to the planning, implementation and evaluation of public health practice, closely integrated with the dissemination of these data to appropriate persons and linked to prevention and

[7] Directive 2003/99/EC of the European Parliament and of the Council of 17 November 2003 on the monitoring of zoonoses and zoonotic agents, amending Council Decision 90/424/EEC and repealing Council Directive 92/117/EEC. OJ L 325, 12.12.2003, p. 31-40.

control (Thacker and Berkelman, 1992). The definitions of these terms are not harmonised across sectors and may be used in a slightly different way. When referring to the integrated analysis across human, animal, feed and food, and processing environment sectors linked to prevention and control of zoonotic infection in the context of 'one health' initiatives (Bidaisee and Macpherson, 2014) the following term shall be used: 'integrated surveillance based on molecular typing for food-borne zoonoses'.

Control actions in animal/food/feed sources to reduce the burden of food-borne illness in the human population (e.g. salmonellosis of egg origin) should be more effectively targeted and evaluated by using the results of integrated analysis of epidemiological and molecular typing data of human and food, animal and feed origin within outbreak investigations and surveillance. In some cases, integrated analysis can also be based on data exclusively from the animal/food/feed sectors but still with the overall aim of prevention of human disease.

The introduction of molecular typing-based surveillance includes the establishment of a continuous information cycle to provide accurate and representative data over time and space, to include the relevant typing characteristics of specified food-borne pathogens (i.e. *Salmonella*, STEC, *L. monocytogenes*, and thermophilic *Campylobacter* spp.) in food animal species and key points in the food production chain. The outputs could then be used for further integrated analyses in combination with corresponding surveillance data from cases of infections in humans. This will help facilitate detection of diffuse outbreaks (Hara-Kudo et al., 2013) and identification and quantification of the sources and transmission pathways for pathogens. The added value of a molecular typing approach to surveillance of food-borne pathogens was strongly supported by the European Commission[8], which in 2012 asked EFSA for technical support regarding the collection of data on molecular typing of food/animal/feed isolates of food-borne pathogens[9]. As a result of this request, it is envisaged that, starting from 2014, the Molecular Surveillance Service (MSS) operated by ECDC will be complemented by a corresponding pilot molecular typing data collection system developed by EFSA, in close cooperation with the European Union Reference Laboratories (EU-RLs) for *Salmonella*, *Listeria* and STEC[10], and which will include results from the molecular characterisation of isolates from animals, feed and food[11]. In that context, real-time molecular surveillance for human cases has been established at the European level using harmonised typing methods. Harmonisation of typing methods for the monitoring of bacteria from food, feed and animals with equivalent methods used in public health surveillance is a priority (ECDC, 2013b). Since November 2012, ECDC has launched the piloting of a new MSS module as part of the European Surveillance System (TESSy) which was successfully evaluated in 2014. The MSS allows Member State (MS) public health laboratories to upload standardised, quality-controlled molecular typing data from clinical isolates of *Salmonella*, *L monocytogenes* and STEC, together with a minimum set of epidemiological data into a EU-shared database (van Walle, 2013). Molecular typing of human isolates is not usually part of routine public health surveillance across the EU. Molecular typing of *Campylobacter jejuni/coli* has been shown to contribute to outbreak investigations (Sails et al., 2003) and to enhance source attribution studies (Muellner et al., 2013; Smid et al., 2013) for campylobacteriosis.

[8] See vision paper from the European Commission on the development of data bases for molecular testing of food-borne pathogens in view of outbreak preparedness available at: http://ec.europa.eu/food/food/biosafety/salmonella/docs/vision-paper_en.pdf

[9] For further details on the request to EFSA for scientific and technical assistance visit: http://registerofquestions.efsa.europa.eu/roqFrontend/questionLoader?question=EFSA-Q-2013-00250

[10] Throughout this Opinion, the term Shiga toxin-producing *E. coli* (STEC), which is also known as Verocytotoxin-producing *Escherichia coli* (VTEC), has been used. It should be noted that the designation for the respective European Union Reference Laboratory (EU-RL) is EU-RL for *E. coli*, including Verotoxigenic *E. coli* (VTEC).

[11] This molecular typing data collection system may later be extended to include other food-borne pathogens such as thermophilic *Campylobacter* upon agreement between EFSA, ECDC, the relevant EU-RL and the European Commission.

At the EU level, such an integrated multidisciplinary approach to surveillance of food-borne pathogens was endorsed by Directive 99/2003/EC and Decision 1082/2013/EU[12], which provides criteria for data collection from humans and food as well as in animal and feed sectors. The need for a strong link between data from public health, animal health and food safety laboratories, including robust epidemiological data, as well as close cooperation between Member States and EFSA, was highlighted by ECDC as an important part of a long-term strategy for the surveillance of food-borne diseases (ECDC, 2013a, 2013b). This requires a high level of interoperability and data integration with other existing monitoring and public health surveillance databases (Figure 1), which is best ensured by prioritising harmonisation issues during the early stages of establishment of the surveillance programme.

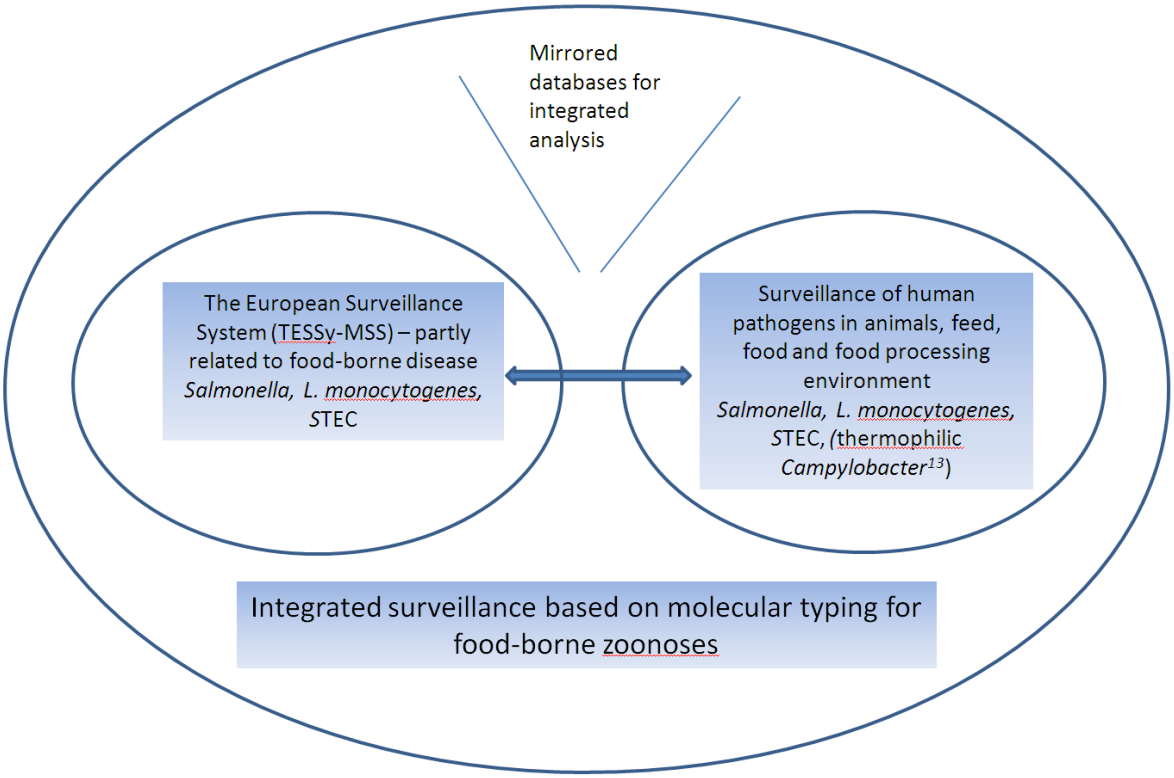

Figure 1: Integrated surveillance for food-borne zoonoses based on molecular typing[13] and its relationship to data managing systems in different sectors

This Opinion sets out to evaluate the requirements for optimising the design of surveillance activities for food-borne pathogens from a molecular epidemiological perspective. In this evaluation, the focus is on the challenges and barriers that exist for surveillance systems to support the requirements of different types of application (outbreak investigation, attribution modelling and the early identification of organisms with epidemic potential) for bacteria within the major food-borne zoonotic groups: *Salmonella*, STEC, *L. monocytogenes* and thermophilic *Campylobacter* spp.. In addition, the challenges associated with the fundamental need for interaction with existing and proposed surveillance systems in the human sector are also considered, which includes a discussion of legal requirements, intellectual property issues, policy for data sharing, and confidentiality. Subsequently, the requirements for a molecular typing-based surveillance system are discussed, including an evaluation of the possibilities for systems based on current control or surveillance programmes for food-borne zoonotic organisms.

[12] Decision No 1082/2013/EU of the European Parliament and of the Council of 22 October 2013 on serious cross-border threats to health and repealing Decision No 2119/98/EC (Text with EEA relevance), OJ L 293, 5.11.2013, p. 1–15.

[13] This molecular typing data collection system may later be extended to include other food-borne pathogens such as thermophilic *Campylobacter* upon agreement between EFSA, ECDC, the relevant EU-RL and the European Commission.

2. Requirements for the design of surveillance activities for pathogens in the food chain employing molecular typing in support of public health surveillance

2.1. Objectives and purposes of molecular typing of food-borne pathogens in animals, feed and food related processing environments by molecular typing as a basis for integrated surveillance

A surveillance programme based on the characterisation of organisms from animals, feed, food and related processing environmental samples should ideally facilitate:

- integrated analysis of molecular typing data, including results of WGS analyses, from different sources and from cases of human infection to support investigations of food-borne outbreaks (i.e. hypothesis generation and confirmation of the animal/food source), source attribution analysis and early identification of emerging food-borne pathogens with epidemic potential;

- description of epidemiological trends in the occurrence of food-borne pathogens (e.g. specific subtypes) in the target animal reservoirs, food, feed and relevant production environments, across geographical regions and time periods;

- description of the pattern of occurrence of specific subtypes among the isolates of a pre-defined pathogen/serovar over regions and time periods;

- detection of unusual epidemiological patterns suggesting the emergence of specific subtypes in pre-defined animal reservoirs and/or foods;

- assessment of the risk of emergence of new subtypes in animal reservoirs and/or the food production chain or established subtypes circulating in unexpected animal species and/or stages of the food production chain.

The requirements for the design of surveillance activities that are addressed in this Opinion will focus on the molecular typing of isolates reviewed in the earlier Opinion in this series (EFSA, 2013a). At present, various molecular typing methods are applied for the subtyping of food-borne pathogens worldwide (EFSA, 2013a). PFGE is still the most widely used method for typing of *Salmonella*, STEC and *L. monocytogenes*. For *S.* Typhimurium and *S.* Enteritidis, PFGE may be used together with Multi-Locus Variable number tandem repeat Analysis (MLVA), although MLVA is increasingly being used as the sole method. These methods are used within surveillance networks such as the PulseNet International and the ECDC-supported Food- and Waterborne Diseases and Zoonoses network (FWD-Net) in the scope of the MSS and also by the EU-RLs and National Reference Laboratories (NRLs) for routine subtyping of *Salmonella*, STEC and *L. monocytogenes* in food and feed, which enables the sharing of typing data as well as epidemiological data among partners. Multi-locus sequence typing (MLST) has been the method of choice for thermophilic *Campylobacter* (EFSA, 2013a), but is now being superseded by WGS.

The increasing availability of rapid and affordable molecular typing tools/methods will lead to progressive modification of the traditional approach to monitoring food-borne pathogens. The increased use of genome sequence-based techniques (EFSA, 2013a) will potentially widen the use of monitoring data beyond their traditional purposes (Segata et al., 2013). As an example, the collection of sequence data from food-borne bacterial genomes is likely to assist with the early identification of organisms with epidemic potential (Sintchenko and Holmes, 2014) as well as contributing to the accuracy of outbreak detection and investigation (Leopold et al., 2014).

Such integrated analyses will be optimised if surveillance activities incorporate complete datasets containing all relevant information on the isolate. Examples of such datasets are those related to the genotype and other characteristics such as serovar or antimicrobial resistance profile, coupled with

accurate data on the effect on the host and related epidemiological data. Obtaining and managing such data represents an important challenge for surveillance networks, together with the need for refining the related bioinformatics analytical framework that is essential for efficient analysis of large quantities of data.

International projects aimed at establishing comprehensive genomic sequence molecular databases of microbial pathogens are currently under development (e.g. Global Microbial Identifier[14] and activities under a 2014 EU-funded research topic within the frame of Horizon 2020[15]). Nevertheless, these prototype databases cannot currently be used for the purposes of surveillance since they are not widely linked to epidemiological data (ECDC, 2013a). Thus, development of linkage mechanisms to access complex genetic and epidemiological data within different databases may be necessary. Bacterial WGS databases that are integrated with epidemiological surveillance databases are now being piloted at national level by public health institutes in Europe and the USA for food-borne disease surveillance and outbreak investigations (Brisse et al., 2014).

Specific requirements for the design of surveillance activities employing molecular typing have been reviewed and optimal features for monitoring of food-borne pathogens in animals, feed and food are proposed, in accordance with legal bases at the EU level (e.g. Directive 2003/99/EC on the monitoring of zoonoses and zoonotic agents, Regulation No. (EC) 2160/2003[16] on the control of *Salmonella* and other specified food-borne zoonotic agents).

2.2. Linkage of molecular typing results with the appropriate level of epidemiological data – harmonised surveillance programmes versus routine laboratory submission

Molecular typing-based surveillance programmes have been particularly helpful in identifying and investigating geographically dispersed common source outbreaks. Molecular typing-based surveillance for food-borne pathogens in animals, feed and food relies on effectively linking molecular characterisation information from the isolates with data on the populations from which the pathogens originated, in order to support integrated comparative analysis with corresponding data from pathogens from cases of human infection (Figure 2). Surveillance based on molecular typing should ideally be based on harmonised monitoring programmes at different production levels (i.e. food animal sectors, food, animal feed or food processing environments) but can also utilise isolates from non-harmonised sampling processes, routine submissions to laboratories or isolates obtained from studies that are not statistically based.

[14] For further details on the Global Microbial Identifier initiative visit: http://www.globalmicrobialidentifier.org/

[15] Details on the EU Horizon 2020 work programme topic PHC7-2014 on 'Improving the control of infectious epidemics and food-borne outbreaks through rapid identification of pathogens' are available at: http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/2250-phc-07-2014.html (last accessed on 26 June 2014).

[16] Regulation (EC) No 2160/2003 of the European Parliament and of the Council of 17 November 2003 on the control of salmonella and other specified food-borne zoonotic agents, OJ L 325, 12.12.2003, p. 1-15.

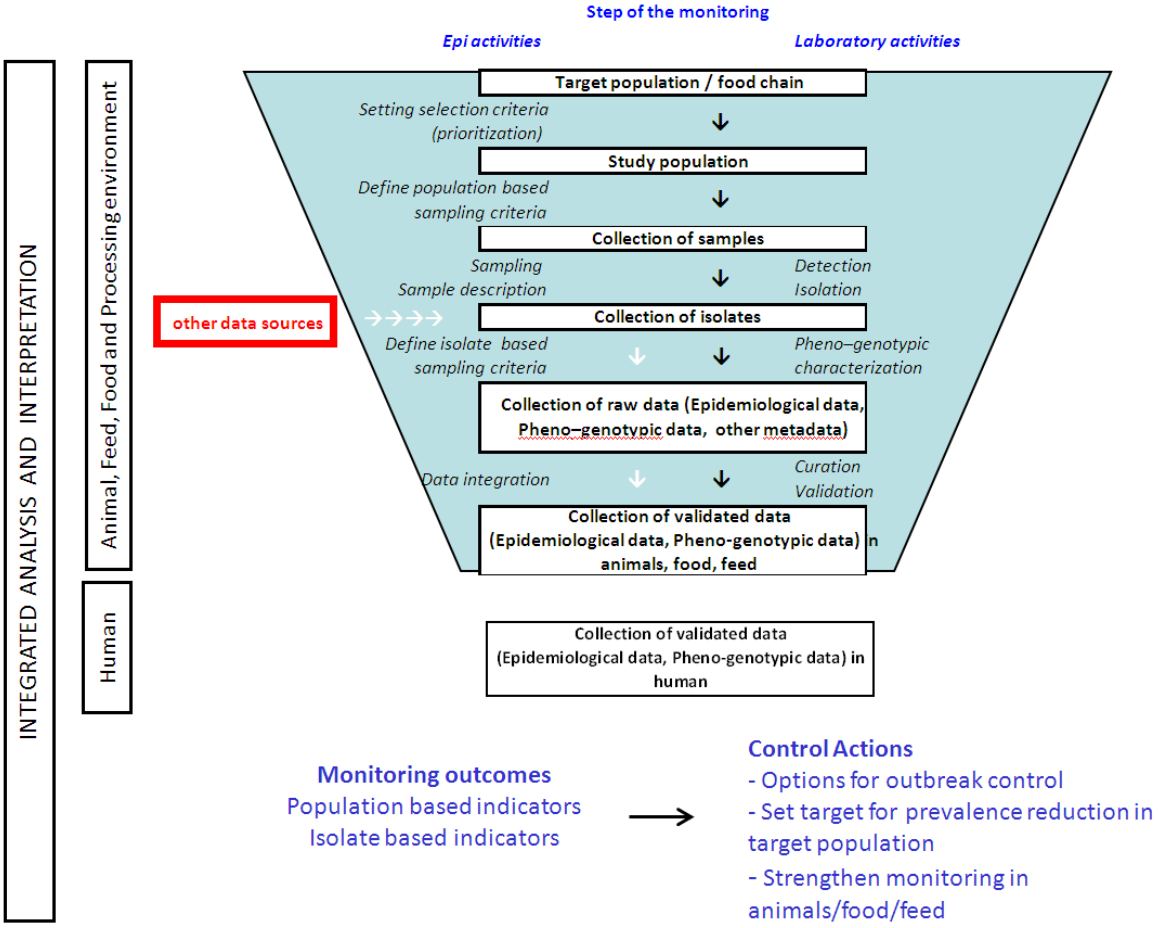

Figure 2: Flow chart of integrated surveillance based on molecular typing for food-borne zoonotic pathogens. Black and white arrows refer to harmonised monitoring programmes sample/data flow and routine laboratory submission data flow, respectively.

The key elements of surveillance based on molecular typing for food-borne zoonotic pathogens in the food and feed sector should be designed *a priori* in close collaboration with the public health sector before the programme is introduced in the field. They can be summarized as follows:

- Target and study populations (e.g. animals/food/feed) should be defined as appropriate for purposes of integrated surveillance and the food-borne pathogen being monitored.

- The sample size and the sampling strategy should be established, taking into account not only the expected prevalence of the pathogen in the target populations/sample sets of animals/food/feed under surveillance, but also the global genetic variability of the food-borne pathogen being monitored.

- The data dictionary for description of the sample, the pathogen and epidemiological information should be provided.

- The optimal discrimination level between subtypes of the food-borne pathogen being monitored in the population under study (see Section 2.4) should be established in agreement with the human sector to support the different applications of the molecular-based surveillance.

- Depending on the specific objectives of the integrated analysis of surveillance data, a minimum set of surveillance outputs to be routinely provided from animals/food and feed should be determined. These should at least include the frequency distribution of the various subtypes of a certain food-borne pathogen isolated in a given population and/or group of animals/food/feed and, whenever possible, prevalence or incidence rates together with the confidence level in the populations/groups under surveillance.

- Once routine data collection is established, threshold levels for frequency/occurrence of food-borne pathogen subtypes in the animal population and food/feed group of interest should be set so that unexpected patterns and trends are identified and action can be taken.

- Joint ECDC-EFSA outbreak assessment procedures should be updated and interpretation criteria should be agreed to enable trace-back of the outbreak source/vehicle based on linkage by molecular typing, following outbreak detection by molecular typing integrated EU public health surveillance of human food-borne infections.

Surveillance programmes based on active and harmonised sampling are the most suitable for statistical analyses and may be used for testing hypotheses. They provide the most complete, accurate and representative data and are more likely to be suitable for source attribution and detailed/advanced epidemiological investigations and risk assessments, as long as the datasets are sufficiently large to support robust statistical analyses. The object of sampling (i.e. the population under study) is a specific animal population and/or food category. The primary sampling unit, which is the starting point for epidemiological data collection, should be precisely defined (e.g. a poultry flock in a defined production sector, e.g. parent breeder for laying hens, a specific ready-to-eat foodstuff). The set of epidemiological data describing the sample unit (e.g. date and place of sampling, type of sample and origin of sample such as animal/food/feed) can be coupled with information on molecular typing characteristics of the pathogen being monitored, but only if a bacterial isolate can be obtained from the sample. Nevertheless, a minimum set of epidemiological data from all the sampling units included in the surveillance programme should, ideally, always be available following sample collection and be independent of the detection and isolation of the pathogen in question. The availability of these data, together with the information on the population under study, the sampling design and the time framework during which sampling was carried out, would support the estimation of a variety of population/source-based analytical outputs and the relative level of the associated uncertainty.

This approach would therefore allow the evaluation of the external validity of a molecular surveillance programme and also the calibration of the, methods used, particularly with regard to their optimal resolution, accuracy and precision Conversely, active, harmonised programmes may have limited flexibility. This is because this approach requires careful definition, *a priori*, of all the main components of the surveillance system. Retrospectively, obtaining additional data, such as that which may be needed for outbreak investigation, may therefore be difficult, although the available data may act as a guide for directing more detailed sampling investigations.

Typing results of isolates collected from routine laboratory submissions where the isolates are linked to limited information can still be valuable and may help support food-borne outbreak investigations, generation of hypotheses, early detection of emerging pathogen subtypes and genetic studies of bacterial populations (David et al., 2013). In a passive monitoring approach, the starting point for epidemiological information collection is represented by the single isolate which is submitted to the laboratory for further molecular typing. In this case, the information on the molecular typing characterisation can be easily obtained, while descriptive data on the original sample unit and source population (i.e. specific animal/food/feed sector), as well as details of the sampling process, may be difficult to obtain or incomplete. Thus, the only analytical outcome that can be obtained via passive monitoring programmes may be an 'isolate-based result', such as the frequency distribution of a certain subtype within a given bacterial group/subgroup. This is an important limitation as the possibility to make inferences relating to the characteristics of animal/food/feed sources may be

limited. However the isolate-based data obtained may still be important, if strains with new risk characteristics are identified, and this may act as a trigger for further structured investigations.

The validity of the integrated analyses depends on correctly linking the molecular typing results to the source and to analyse the data accordingly to the sampling frame. When approaching molecular typing surveillance, distinguishing between external validity and internal validity is important (Dohoo et al., 2009). While external validity relates to the capacity to extrapolate the estimates to the animal/food/feed population under study, internal validity relates only to accuracy of the molecular typing methodology and interpretation in terms of designation of subtypes. This is particularly important as surveillance in some animal, feed and food sectors is carried out *via* structured or harmonised surveillance programmes and the availability of information on the sampling criteria and process affects the possibility to estimate the external validity of the results of the surveillance programme. While external validity is fundamental for source attribution studies, this might not be so important for other applications of molecular typing surveillance.

2.3. Interaction between the discriminatory power of the molecular typing and the epidemiological concordance of grouping isolates into subtypes relevant for the different types of applications

All bacteria are subject to significant ongoing genetic change. Evolution driven by genetic mutation and selection has given rise to highly adaptable organisms that are able to exploit various conditions (e.g. nutrient sources, temperature zones) to expand into novel niches and extend their host range. Such evolution may also be linked to the emergence of various 'epidemic' strains of pathogens, such as certain subtypes of *Salmonella*, in combination with other biological factors and epidemiological opportunities for dissemination (EFSA, 2013a).

When isolates of *Salmonella*, STEC, *L. monocytogenes* and thermophilic *Campylobacter* spp. obtained from surveillance of animals, feed and food are characterized by molecular typing methods, data for measuring the variations in the core as well as in the accessory genome can be obtained. This variability can form the basis for analysis of genetic relatedness and evolutionary relationships among the isolates. A key point for such analysis in relation to integrated public health surveillance is to determine a threshold value for the level of genetic variation amongst isolates that can still be regarded as epidemiologically related. This threshold of epidemiologically relevant genetic/genomic relatedness will vary according to the time frame, population size and geographical scope of the investigation of the chain of transmission (Struelens et al., 1998).

No general rules for determination of the optimal similarity threshold and optimal genomic target regions can be provided as this will depend on the actual bacterium of interest and its genetic nature, for example clonality, genetic stability and the occurrence of horizontal gene transfer (EFSA, 2013a) in relation to the specific type of application. These factors vary in relation to the application purpose for the molecular typing and the requirement to minimize misclassification bias. In each situation there is a need for evaluation of the resolution of typing of isolates in a specific epidemiological context. Validation of these thresholds requires observational studies that include well-documented epidemiological delineation of transmission networks.

The discriminatory power of a method describes its capacity to assign different genomic subtypes to epidemiologically unrelated strains in the population studied (Hallin et al., 2012) and is thereby a tool for describing the threshold for separation of epidemiologically related and unrelated isolates. A high discriminatory power will normally lead to the division of panels of isolates into many subtypes, and the probability of categorizing unrelated isolates to the same subtype will be small. With increasing discriminatory power, the probability of assigning related isolates to different subtypes also increases. In contrast, a relatively low discriminatory power will result in fewer subtypes, where the probability of categorizing related isolates to different subtypes is small, but the probability of including unrelated isolates in the defined subtype increases. There is therefore a compromise between grouping all truly

epidemiologically related isolates with the risk of missing some variants, and grouping isolates with the risk of including some that are not epidemiologically related.

For most molecular methods, it is often possible, as part of the data analysis step, to adjust the discriminatory power of the typing methods by, for example, altering the number of band/loci differences accepted (PFGE/MLVA) or by altering the level of similarities applied for cluster separation. In the integrated analysis of typing data and epidemiological data, it is important to balance the discriminatory power/threshold for separation in a way which gives the most meaningful grouping of isolates from an epidemiological perspective to obtain the highest level of epidemiological concordance (Struelens, 1996).

2.4. Other performance parameters of molecular-based surveillance for food-borne pathogens

When designing an ideal surveillance programme based on molecular typing in animals, feed and food, a set of performance criteria should be followed. These criteria may also be used to evaluate possible alternative programmes in terms of their utility, efficiency and effectiveness. They apply to the surveillance programme as a whole and are defined in relation to a molecular typing-based surveillance approach as follows:

- **Sensitivity:** *Sensitivity* of a surveillance programme is defined as the likelihood that the events under surveillance are identified when they occur (Novick et al., 2008). *Sensitivity* also refers to the ability of the programme to detect outbreaks of disease, including the ability to monitor changes in the number of cases over time (Salman et al., 2003). Thus, for molecular-based monitoring in animals, feed and food, a sensitive programme is characterized by the ability to detect in a timely way changes, over time and space, in the genetic pattern of a bacterial population in a given animal/food/feed sources, such as the emergence of a new or unexpected subtype in a specific animal species. *Sensitivity* is highly dependent on the intensity of surveillance in terms of the number of samples tested.

- **Representativeness:** A surveillance system that is representative accurately describes the pattern of occurrence of a health-related event over time and space and its distribution in the population under surveillance (Buehler et al., 2004). In the case of molecular typing-based surveillance in the context of the current Opinion, *representativeness* can be defined as the extent to which the findings of surveillance accurately reflect trends in specific food-borne pathogen lineage/subtypes of *Salmonella*, *E. coli*, *L. monocytogenes* and thermophilic *Campylobacter* in a defined animal population/food/feed source in a specific area or during a defined period of time. *Representativeness* depends primarily on the efficiency of the sampling process in the animal population or in the food production chain in terms of randomised sampling. Thus, sample-based surveillance programmes relying on harmonised active sampling are preferable.

- **Data quality:** Depending on its various applications, a molecular-based surveillance programme should define minimum acceptable standards for sample and isolate description, including molecular typing data and epidemiological information. *Completeness* describes to what extent the required data are actually available, reliable and representative.

- **Accuracy and Precision:** Statistical *accuracy* indicates how close calculated estimates (e.g. prevalence estimates) based on collected data are to the true value, whereas *precision* indicates how uncertain the calculated estimates are (i.e. how broad the confidence limits are) (Rothman and Greenland, 2008). Both high *accuracy* and high *precision* should be aimed for, but, in general, high *accuracy* is to be preferred, since there is little value in being very confident in a prevalence estimate of a certain subtype in a certain food source, if the estimate is not representative of the sampled population. The Simpson Index of Diversity (D) is an overall measure of subtype distribution that allows for comparison of population diversity over time

or between different populations (Hunter and Gaston, 1988). However, the level of *precision* (and therefore the required sample size) should be defined *a priori*, in order to obtain meaningful and useful estimates. In monitoring programmes, the sample size is often a limiting factor, and the application of molecular typing methods with high discriminatory power may lead to assigning limited numbers of isolates to the different subtypes that have been identified, which may result in limited statistical power in subsequent analyses.

- **Timeliness:** *Timeliness* reflects the duration of time intervals between event occurrences (e.g. the increase of incidence of a new pathogen subtype in a certain category of food), event reporting and information analysis, interpretation and dissemination (Novick et al., 2008). Owing to the high numbers of operational steps included in a molecular-based surveillance programme, *timeliness* may be challenging as the reporting delay may be considerable. *Timeliness* influences the opportunity for public health and food authorities to rapidly take appropriate control initiatives. Any unnecessary delay in the collection, management, analysis, interpretation or dissemination of data for surveillance may affect the ability to initiate prompt intervention or provide timely feedback (Buehler et al., 2004).

- **Flexibility:** *Flexibility* refers to the ability of the typing method used for surveillance to accommodate changes in operating conditions, objectives and/or information needs with little additional cost in time, personnel or funds. *Flexibility* might include the ability of an information system, whose data are used for surveillance of a particular health condition, to be used for surveillance of a new health problem (Buehler et al., 2004). This aspect is particularly relevant for molecular typing surveillance as it influences the possibility to exploit other existing surveillance programmes and to include new typing methods.

- **Stability:** *Stability* refers to the reliability of the typing methods for obtaining and managing surveillance data and to the availability of those data (Buehler et al., 2004). With regard to molecular typing surveillance in animal/food/feed, *stability* is importantly influenced by the consistency of descriptors used for animal/food/feed and sector definition, the descriptions used over time and space as well as the reproducibility of typing results, nomenclature and strain definition.

- **Simplicity and Acceptability:** *Simplicity* refers to the system in operation being easy to use for persons participating in all stages of a surveillance programme (Novick et al., 2008). *Acceptability* reflects the willingness of participants in a surveillance system to fully accomplish the procedures and supply the required data. Both *simplicity* and *acceptability* may be influenced substantially by the time, effort and implicit difficulties required to complete and submit the information or perform other surveillance tasks.

Data requirements depend on the applications for which data are collected and analysed. Although many of the requirements underlying the different purposes are often similar (e.g. accuracy, precision, representativeness), others may differ or even be conflicting, particularly when it comes to conducting the surveillance in practice. As an example, the need for *completeness* of the information might be detrimental for *timeliness*, which in turn is of the highest importance for outbreak investigation. Surveillance programmes should therefore be flexible and balance the different needs and requirements. This could, for instance, mean that molecular typing data should be entered more or less in real time, whereas other data (e.g. epidemiological data) that complete the dataset can often be provided at a later stage.

When building up a molecular typing surveillance programme, although the ideal level of *representativeness*, *timeliness*, *accuracy*, etc. should be satisfied by the programme as a whole, the activities underlying optimal requirements are implemented at each distinct stage of the data collection process only. Difficulties in the coordination of the various steps of the data collection process may exist and importantly affect the overall performance of molecular-based surveillance. For this reason

efforts should be made to remove as much as possible *a priori* any possible barriers between the different actors playing a role in the data collection process.

In summary, depending on its main application, a surveillance system for food-borne pathogens based on molecular typing should be designed to optimise the following components:

- description and comparison of epidemiological trends of occurrence of specific genetic variants over time and space, in different animal reservoirs and food categories: *representativeness, stability, accuracy*;
- hypothesis generation within outbreak investigations and source attribution modelling studies: *sensitivity*, *representativeness, accuracy, timeliness* and *completeness*;
- early detection of emerging epidemiological events such as an emerging subtype in the food chain: *sensitivity, timeliness, flexibility and completeness;*
- studies aimed at identifying genomic markers for newly-emerging genetic variants with the potential for future epidemic spread: *sensitivity*, *completeness*, *representativeness* and accuracy.

2.5. Estimation of statistically representative group of isolates to be included in a molecular typing-based monitoring programme for zoonotic hazards in animals and food

A careful estimation of the sample size and the sampling design necessary to support the various applications of molecular typing surveillance for food-borne pathogens should be established and guide the design of the monitoring activities in animals, feed and food. This *a priori* evaluation would substantially affect the effectiveness of a programme in accurately depicting the epidemiological pattern of occurrence, over time and space, of bacterial subtypes in the various populations and sources being investigated, and in testing hypotheses of association between specific pathogen subtypes and animal sources. As these applications are necessary to support source attribution studies, any statistical consideration on the sampling requirements of a molecular-based monitoring programme in terms of representativeness, accuracy and statistical power seems to be of critical importance for the attribution modelling application of the molecular-based surveillance. In the near future, these will probably become more relevant in establishing the ideal sampling in support of the early identification of organisms with epidemic potential in the various animal population and food/feed sources, whenever the application of WGS will be routinely applied.

Representativeness and accuracy are closely related to the sampling design. A simple randomized sampling is often preferred, as this should ensure that the molecular typing characteristics of the isolates being examined reflect those circulating in the animal population or food/feed sources under study, allowing for conclusions to be drawn for the entire population. However, securing randomization is often not straightforward, due to the structure of the population/source under study. Thorough knowledge about the population to be sampled is therefore essential when designing the sampling plan. As an example, the hierarchical structure of the animal population implies the need to establish criteria for sample stratification among production sectors, between flocks or herds and within flocks/herds. In addition, seasonal (e.g. Christmas turkey or spring lamb production) or geographical variations resulting, for example, from specific food demands and availability, may also need to be considered. Therefore, to achieve representativeness, it is necessary to consider where and how to sample.

Estimation of optimal sample size and sampling scheme can more easily be established whenever animal populations are the object of monitoring, rather than food. In the case of animals, the study population can be more clearly and univocally defined (i.e. in terms of species, age category, production type, etc.) whereas difficulties in establishing the optimal requirements for sampling in food arise from the huge diversity that characterise the different food categories and sources. As a

result, a lack of comparability of results over time and space and/or in the robustness of the estimated trends is more likely to affect surveillance outcomes in foods than in animals.

Precision is another attribute which is closely related to the sample size. Basically, to determine the minimum number of samples/isolates of food-borne pathogens to be included in a molecular typing-based monitoring programme, the following should be taken into account:

- the expected prevalence of the subtype(s) of interest in the various animal/food/feed populations over a time period;
- the desired level of confidence; and
- the accuracy (the error level).

These elements should preferably be set *a priori*, which is a challenge when applying molecular typing methods because their values (particularly the expected prevalence) may vary substantially, depending on the discriminatory power provided by the different typing method(s).

The prevalence estimate of a specific subtype of interest circulating in the animal/feed/food population depends on how the subtype is defined (i.e. by the molecular typing method applied) and which population the estimate may be referred to. Obtaining population-based estimates with the desired level of confidence and accuracy is therefore reliant on a clear definition of the population under study and the chosen level of molecular discrimination. For an EU monitoring programme, the sample size will also depend on the required level of confidence and accuracy, at the EU, Member State or regional level. The more detailed the data requirement, the larger the sample size required.

To estimate the prevalence of a specific food-borne pathogen subtype circulating in a certain animal population or food/feed source, the minimum necessary number of isolates can be calculated using formulae for sampling size estimation of proportion from a population, such as those proposed by Dohoo et al. (2009) or Noordhuizen et al. (2001).

Even though the actual sample size may be simple to calculate, the population under study and the unit to sample may be more difficult to define, particularly for food and feed products. Firstly, it is necessary to consider the level of detail with which food/feed products should be defined (e.g. pork *vs.* ready-to-eat pork *vs.* ham). Secondly, it is important to determine which sample unit is most appropriate, for example batches or amount (e.g. tonnes, kilograms) produced, sold or consumed. Thus, in an active sampling programme, thorough knowledge of the animal population or food/feed source under study and the method(s) for molecular typing of food-borne pathogen of interest should be established. This is required to inform decisions on defining and prioritizing the required level of discrimination necessary to study the different pathogen/source combinations of interest. Ongoing monitoring of trends relating to contamination of the food chain can also assist with detection of clusters of isolates from cases of human infection that might otherwise be missed because of a lack of epidemiological evidence of relatedness. In such cases, the larger the proportion of isolates from cases of human infection that are included in cross-sector analyses, the greater the likelihood of detecting these related strains.

Similar considerations apply to the use of molecular typing for other purposes. In the case of outbreak investigation it is likely that there is already a sub-population of suspect isolates that have been defined by current phenotypic methods that can be investigated in more detail, and in these cases all relevant isolates are likely to be included.

In the case of attribution modelling, representativeness of isolates is the most important factor, and it is difficult to predict the proportion of populations of isolates from different sources that need to be typed in advance without some knowledge of the existing number and distribution of subtypes. A pilot study may therefore be required if there are few pre-existing data.

The re-emergence of 'new' food borne pathogens is regarded as a rare event. To intervene effectively the first cases should be identified before control measures become uneconomic or impractical to be applied.

2.6. Alternative approaches for a molecular typing-based surveillance programme for zoonotic hazards in animals and food

In the previous sections, the focus has been on the requirements for the design of a surveillance programme based on molecular typing in animals, feed and food. Establishment of such a surveillance system will entail several challenges and barriers, as described above. A possible and reliable approach to overcome limitations, difficulties and costs of implementing new activities for optimal surveillance is to collect data and obtain isolates to be further characterized by means of molecular typing, from the already established harmonized surveillance programmes and other activities for surveillance and control of bacterial food-borne pathogens. The use of the experts within existing networks (e.g. EU-RLs and NRLs for food and feed, ECDC-supported EU-Network of NRLs for Public Health Surveillance) as curators of the data generated is also strongly advised by the EU Commission[17] and is under development in a pilot project developed by EFSA[18]. The implementation of databases for the purposes of molecular surveillance may be achieved by complementing the information collected within the various existing programmes (e.g. harmonised surveillance programmes) with the relevant molecular typing data.

The use of other data and samples/isolates collected in the framework of already established harmonized surveillance programmes, as well as other programmes, implies the need to preliminarily assess the utility, efficiency and effectiveness of existing programmes for the purposes of molecular typing and integrated surveillance. This evaluation should reply to issues discussed in Sections 2.3 and 2.4 such as:

i. What level of precision and/or representativeness the use of a certain data/isolate sources would allow?

ii. Which are the populations being monitored?

iii. How complete are the data?

iv. What is the sampling frame for data collection?

In this way, fitness for purpose, limitations, bias and possibilities for integration with other programmes should be critically evaluated and described in order to support both the analyses of the data and their interpretation as well as identification of data gaps. Thus, the possibility to use strains and data sources collected primarily for other purposes should based on:

- a full evaluation of the study design (target population, sampling design, etc.), limits of each programme/survey;

- agreement of what would be ideal: minimum typing programme to be carried out and its representativeness.

A full evaluation of the extent to which each alternative surveillance programme/survey could possibly introduce bias in the molecular surveillance for food-borne pathogens and affect the final

[17] See vision paper from the European Commission on the development of data bases for molecular testing of food-borne pathogens in view of outbreak preparedness available at: http://ec.europa.eu/food/food/biosafety/salmonella/docs/vision-paper_en.pdf

[18] For further details on the request to EFSA for scientific and technical assistance visit: http://registerofquestions.efsa.europa.eu/roqFrontend/questionLoader?question=EFSA-Q-2013-00250

outcomes of the analyses (with regard to the different question addressed) is required, particularly regarding the application of different WGS methodologies and data analysis pipelines.

Sampling activities to detect *Salmonella, L. monocytogenes*, STEC and thermophilic *Campylobacter* in animals, feed and food are routinely but variably undertaken under the auspices of many different programmes and with different aims, at both the EU and the Member State level.

Established surveillance programmes, such as those implemented in the framework of the mandatory *Salmonella* control programmes in animal populations (Carrique-Mas et al., 2008; EFSA, 2009) or harmonized surveys such as the baseline studies (Regulation (EC) No 2160/2003), may be considered the best options to allow estimation of population/source-based factors that are comparable among countries and between years and seasons. Other alternatives are less suitable for this. These include sampling activities related to official control to evaluate the compliance with microbiological criteria (for *Salmonella* and *L. monocytogenes* only), control programmes organised by industry and Hazard Analysis and Critical Control Points (HACCP) sampling plans at the food business operator level, as well as results from clinical investigations, specified suspect sampling and sampling for outbreak investigations.

The usefulness of the alternative data and isolate sources for the purposes of molecular-based surveillance varies dramatically. Difficulties in the interpretation of the alternative data and sources of isolates are primarily connected with the complexity of gathering the relevant epidemiological data to be coupled with the molecular typing data and clearly defining/identifying: (i) the population under study; (ii) the sampling stage; (iii) the sampling design/criteria; (iv) the unit of sampling (i.e. single sample or batch) and (v) the rule for sampling and testing of isolates.

Based on a general evaluation of the established programmes and various sources of isolates and data, it is possible to summarize the currently available data sources:

A. EU-wide continuous surveillance

In the animal populations, isolates and epidemiological data from EU-wide continuous and harmonized surveillance programmes are highly valuable information sources. Harmonized programmes are only available for *Salmonella* in poultry (breeding hens, laying hens, broilers and turkeys) (Regulation (EC) No 2160/2003) where *Salmonella* is isolated from well-defined populations based on harmonised sampling plans and laboratory methods in all Member States. This will give a high level of representativeness and completeness, but flexibility is low. Isolates obtained from within these programmes could be collected for further molecular characterisation and typing results could be coupled to the existing epidemiological data and form the basis for extended analysis, including integrated cross-sectoral analysis.

It is likely that new requirements for harmonised collection of isolates from slaughtered animals for the purposes of monitoring antimicrobial resistance in food animal populations will offer new opportunities for typing targeted bacterial species (e.g. thermophilic *Campylobacter*, or recovering additional bacteria from the test samples).

B. EU-wide baseline studies (of greater value if current)

Several baseline studies have been carried out during the last decade[19]. These studies generated collections of isolates based on harmonized sampling and microbiological analysis. In addition, these studies have shown that the sensitivity of the baseline studies was higher than for the subsequent harmonized surveillance (EFSA, 2007, 2008). Isolates obtained from these studies may be available for further molecular characterisation and typing (if they are stored by the laboratories) and may offer high sensitivity, representativeness for the time of the study and completeness. This source of

[19] http://www.efsa.europa.eu/en/zoonosesscdocs/zoonosessurvey.htm

information has no flexibility as the value of these data will decline as the time from the study increases.

C. (Multi-) state/national continuous harmonised surveillance

Other monitoring programmes and cross-sectional surveys undertaken to some extent on a regular basis may be available at either the national or the multi-state level in various animal populations or food commodities. These studies are, in general, reported by EFSA in the annual Zoonoses Summary Reports[20] but compared with EU harmonized programmes these data are usually considered of lower informative value and the findings (e.g. prevalence and distribution of subtypes), even if comparable and representative of the same targeted population, might be influenced by different sampling approaches, testing methods and/or sample description detail across the Member States. As a result, varying levels of accuracy, inclusivity, completeness, precision and discriminatory power of the programmes may then affect the comparability of estimates among countries and limit the possibility for further inference of results at the EU level.

D. National official control programmes to evaluate the compliance with microbiological criteria (Regulation (EC) No 2073/2005[21])

Despite the small number of animal species/sectors or food/pathogen combinations being monitored under the framework of EU harmonized programmes, data from official control plans to assess the compliance of foodstuffs with microbiological criteria (Regulation (EC) No 2073/2005) and from control programmes organised by industry (see below) can, potentially, greatly widen the range of food items and food processing stages being monitored. For *Salmonella*, microbiological criteria are designated for various types of meat products and products thereof, cheese, milk powder, ice cream, eggs, ready-to-eat foodstuffs, cooked crustaceans, live bivalve molluscs, fruit, vegetables and juices. For *L. monocytogenes*, the microbiological criteria concern either the industrial processing or the retail level for products of meat origin, ready-to-eat foods, various types of cheese, dairy and fishery products. For STEC, microbiological criteria concern only sprouted seeds. No microbiological criteria are currently available for thermophilic *Campylobacter* in most countries, but this may be subject to change following recent discussions led by the European Commission, and voluntary monitoring programmes are in place in many countries. Few standardised alternative sampling activities are in place across the EU, limiting the number of alternative isolate and data sources that can also be used for the purposes of harmonised molecular surveillance).

The official control of the industrial compliance with microbiological criteria may be seen as an alternative source of isolates available for further molecular characterisation. However, although specific rules for sampling and testing, as well as standards for sampling unit definition, are available, the criteria for sampling design are usually poorly defined. Representativeness cannot be considered optimal as sampling is usually risk based rather than randomized, and isolates are not required to be phenotyped (e.g. serotyping for *Salmonella* or speciation for thermophilic *Campylobacter*) or stored. As a result, there is a lack of comparability across regions/countries and over time. However, as these surveillance activities are included in the multi-annual National Control Plans (NCPs)[22] implemented by each Member State, these alternative data/isolate sources may provide valid and accurate information at the national level, whenever specific programmes /targets (and criteria for sampling) are implemented.

E. Harmonized industrial investigations (microbiological criteria)

The food industry has to perform its own control investigations in order to document the compliance with the general food safety criteria defined in the legislation (Regulation (EC) No 2073/2005). As

[20] http://www.efsa.europa.eu/en/zoonosesscdocs/zoonosescomsumrep.htm
[21] Commission Regulation (EC) No 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. OJ L 338, 22.12.2005, p. 1-26.
[22] http://ec.europa.eu/food/animal/diseases/index_en.htm

described above, there are specific rules for sampling and testing as well as standards for the sampling unit definition although the representativeness of isolates coming from this source is not satisfactory for surveillance purposes. In addition, the legislation does not oblige the industry to type, store or share the isolates, which belong to the industry, for additional analysis.

F. Routine laboratory submissions

Despite the requirements described for isolates being beneficial for inclusion in a surveillance programme based on molecular typing, all isolates can add some value when included for further molecular characterisation. For investigation of food-borne outbreaks and for early detection of strains with future epidemic potential, timeliness in characterization of isolates is essential and, even though isolates submitted for veterinary diagnostics or for characterisation in relation to control measures implemented by industry are not fully representative of a production sector or a region in general, they may still provide valuable information.

2.7. Additional challenges for molecular typing-based surveillance by applying sequence-based typing methods

On-going and rapid advances in the understanding of the molecular characteristics of bacteria and their genetics, linked to technological developments, will ultimately lead to the use of bacterial WGS methods for food safety applications (EFSA, 2013a). When new WGS methods are applied, additional challenges will emerge. These will be related to the need for sophisticated hardware and software, harmonised data formats and methods for interpretation of data (bioinformatics), ontology and nomenclature (Aarestrup et al., 2012). This highlights the importance of the conclusions stated in first part of this Opinion (EFSA, 2013a) "*cross-sector and international coordination of method validation is required as a priority*" and "*development and improvement of international initiatives with regard to harmonized platforms for sharing of data such as those promoted by PulseNet and ECDC/EFSA should be urgently prioritized, including the integration of WGS into such platforms*".

The challenges of applying WGS are in the process of being discussed. Similarly, management issues and tools are being developed, evaluated and prioritised, together with common understandings and agreements. It is therefore not within the scope of the present Opinion to give specific recommendations on how to design the systems and platforms for WGS when applied in integrated surveillance based on molecular typing for food-borne zoonoses, but to support strong international initiatives for dialogue and agreement on future collaboration in relation to molecular-based surveillance across sectors and countries. An example of such an initiative is EFSA's Scientific Colloquium no 20 on '*Use of Whole Genome Sequencing (WGS) of food-borne pathogens for public health protection'*, which was held in Parma, Italy on 16-17 June 2014.

WGS typing relies on access to sophisticated hardware and software, and the creation of large and complex molecular databases. For food-borne pathogens, the results should be underpinned by skilled interpretation of data and cross-referenced to characterisation from conventional technologies, background data related to the processes carried out and epidemiological data relating to the sample. Laboratory Information Management Systems (LIMS) should therefore be linked to databases containing DNA sequence data and the results of further analysis of these data, by mechanisms such as integrated rule-operated data systems. This is vital for efficiency and to minimise the potential for data re-entry errors. These systems currently require skilled IT system administrators and a common programming language must be integrated into all shared data networks. Web-based systems (e.g. Galaxy and similar software) allow integration of multiple analysis tools, thereby facilitating both harmonised and customised approaches to further analysis of data. Data storage capacity and computing power for rapid analysis are potential limiting factors, and it has been suggested that only raw DNA data should be stored long term, assuming that these data can be rapidly re-analysed when needed at costs lower than those resulting from data storage. If this is done, robust audit trails are essential as there may be variations in analytical resulting from software- or operator-related factors.

Validated information for public health action and intervention requires reliable integration of organism-specific knowledge with genetic code data. This should help identify key associations between aspects of genetic sequences (genotype) and the organism's observable characteristics (phenotype), such as antibiotic resistance profile or virulence attributes, which may be controlled by epigenetic factors not readily available by sequence analysis alone.

The application of genomics to the characterisation and surveillance of food-borne pathogens will have an absolute requirement for rapid, robust and accurate interpretation of data. To interpret properly the wealth of information delivered by WGS, expert bioinformatics support is essential, although the rapid development of more user-friendly analytical software is likely to reduce the need for such specialism. Furthermore, while the requirement for phenotypic identification may be reduced in the foreseeable future, it currently remains an essential part of organism characterisation.

Central coordination and curation is required to ensure the quality both of the collection of molecular data and of the laboratory and epidemiological techniques used to build large evidence bases for food-borne pathogens. Of paramount importance is that the legal framework for the central coordination and curation is in place before any large collection of data is initiated. Substantive variations are already occurring in the systems and approaches used by different organisations for production and analysis of DNA sequence data, since expert programmers are able to combine multiple software options and write bespoke computer programs that suit their immediate local needs and are continually updated, but these are not necessarily easy to harmonise internationally. Suitable reference strains, DNA and sequence data should be made available to investigate the key stages in DNA extraction, primary sequence determination and sequence interpretation. Some comparisons of free-access and chargeable sequence data interpretation pipelines have already been carried out and have shown a reassuring level of concordance, although the level of detail of the gene information identified by different systems may vary (Anjum and Thomson, 2013). Other relevant considerations will include changes to the analysis of epidemiological data, which will require shared databases and secure data transmission to facilitate the required complex interpretation, whilst at the same time maintaining the levels of confidentiality required by different stakeholders.

The principles and specifications for selection of isolates for long-term archiving should also be considered. WGS will be very helpful for selection of strains that are representative of the genetic variation that is present in the strain population, but the origin of strains should also be taken into account in the selection process. Storage conditions for strains should aim to limit further genetic and phenotypic change during storage as much as possible. Currently, freezing on cryobeads or in cryoprotectant media at -80 ^{o}C or lyophilisation are the most suitable options for most organisms, but large collections are expensive to maintain in frozen storage and lyophilisation requires expensive specialist equipment and may result in damage to sensitive micro-organisms.

2.7.1. Outbreak identification

Utilising genomic methods to recognise outbreaks of food-borne pathogens and to assist in the identification of newly emerging organisms with epidemic potential will need suitable high quality bioinformatics together with internationally accessible databases. Such databases will need to be harmonised, well managed, in widespread use and regularly updated with contemporary data submitted in a timely manner, and subject to international accreditation to ensure the production of similar data, regardless of the detailed methodology used to generate the data. Furthermore, inclusion of molecular data into existing EU-wide databases for food-borne pathogens will require strict enforcement of agreed EU standards to ensure that appropriate phenotypic and epidemiological components are also captured. As with phenotypic data, such standards will have to be strictly defined, together with a quality framework, to enable the early and accurate detection of emerging microbial transmission patterns and of organisms with newly emerging virulence characteristics. Particularly important is the need to avoid incorporating misleading data, which may in turn give rise to incorrect conclusions resulting in inappropriate actions at local, national and international levels. Data

dictionaries and data validation and cleaning procedures used to optimise the value of shared international databases must therefore be carefully designed and agreed from the outset.

2.8. Concluding remarks for the design of surveillance activities for pathogens in the food chain employing molecular typing in support of public health surveillance

Strategies for collection of data from molecular typing-based surveillance of food-borne pathogens in animals, feed and food should:

- be based on active sampling and a robust statistically based sampling design should be prioritized for the purposes of molecular surveillance of food-borne pathogens in animal, feed and food. The possible use of alternative sources of strains and data should be carefully evaluated according to the required outcome and to a set of established criteria;
- apply appropriate molecular typing methods and interpretation criteria based on both the pathogen to be characterised and the level of discriminatory power required, depending on the required application of the surveillance results;
- ensure the optimisation of programmes for the use of WGS for molecular surveillance of food-borne pathogens as an urgent priority.
- ensure that molecular typing data are coupled with a minimum required set of epidemiological data (e.g. including information on the sampling context and population/sample set under study);
- ensure that information on the total number of samples that are analysed under EU-wide harmonised surveillance programmes (denominator data), and not just positive samples, is included;
- ensure that datasets are comparable and suitable for joint analyses with other datasets from parallel surveillance in humans and/or other relevant sources;
- optimize activities and efficiency of surveillance by adopting criteria for the identification and the prioritization of the relevant combinations of pathogens /animal or pathogen/food;
- ensure that surveillance activities are primarily aimed at investigating the priority hazard/source combinations and are robust and statistically-based;
- ensure that rules for assembling strain collections and associated provenance data from general surveillance of pathogens are agreed and introduced as EU standards.

3. Requirements for integrated and harmonised data collection and management in relation to molecular typing

Expanding the existing surveillance activities in the animal/food/feed sector by adding molecular typing for further resolution of food-borne pathogens provides new challenges. Some difficulties are already recognized where cross-sector networks are established using surveillance based on typing methods such as MLST, PFGE and MLVA. The main challenges relate to harmonization and sharing of data between the human/veterinary/food and feed sectors. This includes typing data as well as epidemiological data, and requires cooperation among sectors/partners and common legal rules (laws/legislation), which could give rise to confidentiality issues. In the following sections, the overall need for international and cross-sector harmonisation of molecular typing will be reviewed.

3.1. General guidelines on data needs

Integrated surveillance can be defined as a data collecting system enabling the combined analysis of data from both humans and other possible sources including food-producing animals, feed and food. Ideally, the samples from each source are collected through harmonised surveillance programmes so that the resulting data can be used to obtain prevalence estimates that are comparable between regions/countries and over time. Isolates from humans, animals and food should be characterised using the same typing methods and, most importantly, typing data should be accompanied by the relevant epidemiological data needed to analyse and interpret these data. Data for integrated surveillance often originate from different registries/databases (e.g. laboratory databases, central husbandry registries or patient registries). Unique identifiers should therefore be agreed and applied so that data from different databases can be merged appropriately. In relation to integrated surveillance based on molecular typing data, this means the need for detailed specification of the subtypes of food-borne pathogens and the methods used.

3.1.1. Data requirements for integrated cross-sectoral surveillance and analysis

For the purpose of EU-wide integrated surveillance for food-borne zoonoses based on molecular typing, including source attribution studies and joint food-borne outbreak assessments, it is essential that basic data from human public health surveillance and surveillance within animal/feed/food production can be combined and shared in common databases. Expanding the information (epidemiological data) already included in existing surveillance programmes to include molecular typing data for further resolution of the food-borne pathogens highlights the need for standardisation of molecular typing data. This would require cross-sectoral agreement on the use of the same (or fully transposable) typing method, standard operating procedures (SOPs) for assignment of standard type nomenclature and type result archival format as performed for typing of animal, feed and food isolates. All typing information, as agreed in a harmonised cross-sectoral surveillance protocol, that is compatible with existing protocols for the human sector should be reported, as far as possible, in a standard format allowing comparison. Genotyping methods should be validated and adequately supported by external quality assessment (EQA) schemes and data curated by expert database managers.

At present, sufficiently validated molecular typing methods exist for PFGE and MLVA and validation of WGS-derived sequence analysis methods is developing rapidly.

The EU-wide cross-sectoral/national agreements should also include agreement on how to share and analyse data and how to present results and conclusions. The possibility of combining human epidemiological data and typing results with the corresponding data from the animal/feed/food sector may result in politically, commercially or ethically-sensitive findings, which cannot be shared due to data protection legislation or national legal rules. Therefore, stakeholders may be reluctant to share information without a clear agreement on confidentiality, intellectual ownership and the use of data.

Overall, there is strong support for extending the sharing of genomic typing data among scientists from different sectors and countries to include representative isolates from sporadic cases, as well as food and food animal sources, and to accelerate the generation of knowledge. This will facilitate well-founded decision-making in the face of an emerging disease or a current public health threat, although in practice there are several barriers to implementation yet to be overcome. An expert meeting held in Utrecht in The Netherlands, from 6 to 8 December 2009, focused on the discussion of the challenges related to sharing the raw sequence data but also highlighted the need for sharing additional information: "*There is a need to combine biological data across geographic and disciplinary boundaries. A sequence by itself is just a sequence. With epidemiological information as simple as a detection date, place, and source it is already much more, and numerous other relevant bits of information can be thought of.*" The experts concluded that "*to clear the way towards a situation where data-sharing for public health purposes is self-evident, a number of problems need to be addressed. These problems are of legal, economic, technical, institutional, scientific and socio-cultural nature*" (Siebenga et al., 2009).

A key issue is how to balance individual/industry interests and needs against those of public health. This dilemma also arose when considering the surveillance of food-borne disease in humans and control of bacteria pathogenic to humans in animals, feed and food. Nevertheless, expanding surveillance activities by adding molecular typing data to the reporting of the food-borne pathogens, strengthens the connection between human disease and the potential source/vehicle. Molecular typing can also strengthen the legal basis for recall of contaminated food by reducing the uncertainty when linking the specific food to human disease (Rump et al., 2013). Agreement is required to ensure that isolates collected for a specific purpose (e.g. surveillance or control), can be used in other contexts e.g. outbreak investigations or legal proceedings.

Cross-sectoral collaboration is crucial in order to ensure appropriate linkage of animal/feed/food data collection with the well-established public health surveillance of food-borne diseases in humans across the EU and European Economic Area (EEA) countries. This surveillance includes systematic and regular data reporting of human cases of specified food-borne diseases to TESSy as well as the recently established isolate-based molecular surveillance of clinical cases (MSS). By combining human data from molecular typing of isolates with data from animal, feed and food isolates, it should be possible to obtain the optimal information required for multi-country outbreak detection, source-hypothesis generation, epidemiological and trace-back investigations as well as source attribution studies.

Incorporation of the potentially varying requirements for systems supporting at least the three different types of applications that are included (outbreak investigation, attribution modelling and the early identification of organisms with epidemic potential) within one common data management system is an important challenge. A possible solution may be to develop a data management system that allows a high level of epidemiological data which can subsequently be linked to the subset of isolates characterized by more detailed typing results including the results of molecular as well as additional phenotypic tests.

The greatest challenge is to ensure confidentiality and appropriate use of these data. This is mostly a matter of trust and legal certainty between stakeholders and sectors. Therefore, clear agreement on data confidentiality, appropriate use of data and respect of intellectual property rights is crucial. Technical solutions exist. Regarding data use and access policies there is a need to ensure harmonisation across sectors. While the veterinary sector is largely managed through EU legislation, EFSA and EC, surveillance in the public health sector is largely based on operational agreements between Member States and ECDC, like the TESSy data use policy, which has been approved by the ECDC Management Board[23]. In principle, cross-sectoral surveillance protocols based on molecular typing could follow the general steps, which have been used to establish human disease surveillance at the EU level:

- Define the microorganisms to be covered by cross-sectoral surveillance (basically fulfilled by the EC in their request to establish molecular typing data collection for *Salmonella, L. monocytogenes*, STEC and possibly thermophilic *Campylobacter*).

- Define the objectives and purposes for surveillance. These objectives are mentioned several times in various publications but it would be worth collecting them all under the same title with justifications.

- Define the minimum datasets that are needed from human and veterinary sectors for achieving the surveillance objectives and purposes.

- Review the existing data collection systems in both sectors and identify attributes, for example the variables and reporting frequencies, that might require cross-sectoral harmonisation and

[23] http://ecdc.europa.eu/en/activities/surveillance/TESSy/Documents/TESSy-Policy-data-submission-access-and-use-of-data-within-TESSy-2011%20revision.pdf

revision to achieve surveillance objectives and purposes. These needs should be further discussed within and across sectors.

- Agree on standard methods for typing, particularly on molecular typing, and ensure the quality of collected data through appropriate data curation.

- Support capacity and competence building by organising cross-sectoral EQA schemes for laboratories, particularly for molecular typing methods, and organise cross-sectoral training on molecular typing techniques.

- Define baseline SOPs for regular data analyses (e.g. cluster analyses, outbreak investigations, source attribution studies) whilst still allowing flexibility for additional development of improved methodologies.

- Ensure integration of cross-sectoral data access and use policies in the sectoral guidance documents (e.g. TESSy data policy and equivalent documentation relating to animal, feed and food production sectors).

A possible approach to achieve all of the above would be to establish a joint EFSA-ECDC-EU-RLs committee for the support of cross-sectoral surveillance. This committee should consist of experts from public health and veterinary sectors, as well as epidemiologists and microbiologists from both sectors to ensure balance and representativeness in expertise. Appropriate EU industry representative bodies may also be considered to enhance collaborative agreements.

3.2. The data collection process and the objective of harmonisation

Molecular typing-based integrated surveillance for food-borne pathogens supports the need to compare trends in human disease with those in animals/food and feed, over time and geographical regions. To do this, the data collection processes and the characteristics of the data repository should ensure the highest level of reliability of data and results over time and space, as well as the compatibility and interoperability among different systems. Unfortunately, the opportunity to exploit molecular typing data collected for a specific limited study for other purposes, and to optimise the cost-effectiveness of molecular surveillance is currently often hindered by the lack of harmonization.

Compared with other surveillance programmes, monitoring based on molecular typing may involve information cycles which include several participants with different levels of expertise, backgrounds and skills. As an example, it should be considered that not only are sampling, laboratory analyses and data collection/storage usually performed by different parties, but also that each of the isolation and typing steps is carried out by different laboratories (i.e. primary testing and reference laboratories). These conditions may complicate data collection and represent a critical issue that can compromise the effectiveness not only of the data collection, but also of the overall surveillance programme.

The key factor in terms of usefulness of the data management system for integrated analysis across sectors is that the included data should be accurate and comparable at the relevant level. This will require intensive endeavours to achieve harmonisation and standardisation across the different stages of the information cycle, including a strong focus on the quality of epidemiological data as well as the laboratory results. The ultimate goal of harmonization is to reduce inaccurate data collection and analytical interpretation and to avoid biased estimation of surveillance indicators, and/or unnecessary sampling and laboratory testing.

Harmonization is also the key prerequisite for dataset integration and is essential to maximise the opportunities that new information technologies make available, in particular the possibility to virtually connect and query in real time the large datasets for integrative translational bioinformatics studies. Full interoperability between molecular typing datasets is also necessary to provide the appropriate background of integrative data sharing on food-borne bacterial genomes to support studies

of bacterial pathogenicity or virulence, which are necessary for supporting the early identification of organisms with epidemic potential. These data sharing options will need to include user access control mechanisms and anonymisation procedures as required, complying with EU legislation on personal data protection.

The implementation of a common harmonized molecular typing programme, at the EU level, and the establishment of detailed informatic standards for data production and collection (see the European Commission mandate to EFSA and ECDC for data collection) would provide Member States with the opportunity to harmonize their own standards for data collection together with public health laboratories.

3.3. Optimal requirements for harmonisation

'Standardization' reflects the extent to which procedures for data collection and analyses meet uniform standard requirements, such as SOPs across the various steps of surveillance. Harmonization reflects the extent to which different procedures result in the same or mutually compatible outcome. Despite this difference, the term harmonization is currently used to indicate both general purposes of standardization and harmonization.

Although the focus of harmonization is mainly directed at methods for detection and molecular typing characterization of isolates and their results, the scope of harmonization goes beyond the analytical phase. Also included are aspects such as the adoption of an unambiguous terminology and unit of reference in the pre-analytical phase for describing the sampling process and its context (the animal or food source), and the criteria used for attributing nomenclature, cluster analyses and the interpretation of results in the post-analytical phase.

3.3.1. Pre-analytical phase

The objective of harmonization in the pre-analytical phase should focus particularly on the description of:

i. the sample collection process;

ii. the sampling context and

iii. the matrix sampled and being analysed.

A careful and harmonized description of the target population, the study population and the sampling criteria is important to meet the necessary requirements of a surveillance programme in terms of epidemiological data. The availability of information on the sampling stage and strategy would also enable the data gathered within an active harmonized sample-based surveillance programme to be clearly identified and linked with the original sampling design.

All these aspects are unambiguously addressed by the EFSA guidance Standard Sample Description (SSD) (EFSA, 2013b) which provides detailed and harmonized reference standards for data collection by way of a multi-level hierarchical descriptive approach and by the adoption of a controlled terminology in the various collection domains, including zoonotic agents in food, feed and animals. It includes lists of standardised data elements and is proposed as a generalised model to harmonise the collection of a wide range of measurements in the area of food safety assessment.

Harmonization in the pre-analytical phase should also deal with the criteria for the sample selection from the study population (e.g. the sampling unit in animals, the matrix sample from food, choice of isolates to be further characterized) as well as with methods for preparation of test samples. Depending on the existence of specific legal or quality assurance programme requirements, animal, feed or food sampling might be carried out according to the available reference International Standards Organization (ISO) standards or SOPs. As a general rule, in the absence of specific standards or

guidelines, sampling should be consistent with other available guidance in the food safety domain such as those issued by EU Reference or the principles of Good Practice referred to in Article 7 of Regulation (EC) No 852/2004[24].

The use of the ISO standards is prescribed by Regulation (EC) No 2073/2005 on microbiological criteria for foodstuffs and in the specific norms regulating the harmonised monitoring and control programmes for *Salmonella*, thermophilic *Campylobacter* and *L. monocytogenes* provided by Regulation (EC) No 2160/2003, which also provides specific guidance for the sampling process in certain animal categories.

3.3.2. Analytical phase: harmonisation within the EU laboratory networks

Monitoring activities of food-borne pathogens should incorporate detailed information on genotyping and phenotypic characteristics. The laboratory assays should ideally be performed using the same harmonised methods, including nomenclature, in both human, animal and food safety domains. The main scope of harmonisation in the analytical phase is therefore to ensure the highest level of reproducibility of the molecular typing characterization of the isolates and compatibility among methods. This may be achieved by:

- development, validation and dissemination of reference analytical methods, materials and standards for detection in relevant sample matrices and typing of the pathogens isolated;
- assessing the application of the reference methods by the laboratories, in particular by organizing EQA proficiency testing programmes and
- organization of training programmes for the laboratories involved in food control to evaluate the analytical performance in applying the standard methods to specific matrices.

In the animal, feed and food sectors, these activities are organised at the EU level by networks of the NRLs for *Salmonella*, thermophilic *Campylobacter*, *E. coli*[25] and *L. monocytogenes*, which have an important role to facilitate harmonisation of methodology among Member States. Each network is coordinated by the correspondent EU-RL which is appointed according to Regulation (EC) No 882/2004[26]. Each NRL collaborates with the EU-RL and promotes the harmonisation at the national level. The activities and the tasks of the EU-RL should be mirrored at national level by the NRLs. The final aim of this cascade system is to harmonise the approach to hazard detection and identification in animals, feed and food across the EU with the expected result that the official controls conducted on any foodstuff are carried out using harmonised methods and with comparable levels of proficiency throughout the EU. Another important added value of laboratory networking within the EU is the possibility for flexible provision of scientific advice to Member States by way of the same cascade mechanism. This may be particularly important whenever new methods for detection and characterization of food-borne pathogens are implemented and quickly disseminated to elicit a harmonised response across the EU. This is a crucial role and responsibility during epidemic outbreaks of food-borne infections such as the international outbreak of *E. coli* O104:H4 in 2011.

Similarly, the role of EU-RLs and NRLs in promoting collaborative studies on both research and monitoring, including proficiency testing, is very important. The recent collaborative molecular typing study 'ELiTE', on *L. monocytogenes* launched by the ECDC in collaboration with the EU-RL for *L. monocytogenes* and entrusted by the ECDC, was an important example of laboratory harmonisation and integration of the human health and food production sectors. The study is a joint collaborative

[24] Regulation (EC) No 852/2004 of the European Parliament and of the Council of 29 April 2004 on the hygiene of foodstuffs. OJ L 139, 30.4.2004, p. 1-54.

[25] Throughout this Opinion the term Shiga toxin-producing *E. coli* (STEC), which is also known as verocytotoxin-producing *E. coli* (VTEC) has been used. It should be noted that the designation for the European Union Reference Laboratory (EU-RL) is EU-RL for *E. coli*, including verotoxigenic *E. coli* (VTEC).

[26] Regulation (EC) No 882/2004 of 29 April 2004 on official controls performed to ensure the verification of compliance with feed and food law, animal health and welfare rules. OJ L 191, 28.5.2004, p. 1-52.

exercise between ECDC, EFSA, the EU-RL for *L. monocytogenes* (EU-RL Lm) and Member States' public health and food safety authorities. The expected result is a joint ECDC-EFSA-EU-RL Lm report on the molecular epidemiology of *L. monocytogenes* infections in 2010-2011. Importantly, the ELiTE project governance includes coordination by a multi-stakeholder Steering Group (ECDC-EFSA-EU-RL Lm) and scientific support by a cross-sector, multi-disciplinary *Listeria* expert study group.

At the EU level, the development and dissemination of reference methods (i.e. SOPs) for the molecular characterization, as well as the implementation of EQA studies, offer important opportunities to achieve the harmonisation necessary between the typing activities performed for human surveillance and monitoring of hazards in animal, feed and food. This should ideally be achieved by:

i. harmonising the molecular typing methods adopted by the Member States in collaboration with ECDC (van Walle, 2013) and the EU-RLs and disseminated through the network of public health laboratories;

ii. performing joint EQA studies (ECDC 2014 PFGE and MLVA EQA reports) and training sessions.

The choice of detection and typing methods should be ideally based on the following order of priority and also take into account the legal requirements in the different food safety domains, when applicable:

i. ISO/CEN (European Committee for Standardization) international standard;

ii. reference method (SOP) developed by the EU-RL and

iii. other validated internal methods.

As ISO/CEN international standards for molecular typing characterization of *Salmonella*, *E. coli*, *L. monocytogenes* and thermophilic *Campylobacter* are not currently available, it is important to mention that, based on the European Commission mandate, the EFSA has recently invited the EU-RLs for *E. coli*, *L. monocytogenes* and *Salmonella* to compare and evaluate different available methods for molecular typing of isolates under their responsibility for outbreak detection and epidemiological surveillance and to develop harmonised SOPs for: (i) PFGE/MLVA testing of isolates from food, feed and animals; (ii) acquisition, normalisation and quality assessment of PFGE profiles/images of isolates from food, feed and animals; and (iii) curation of the molecular typing data on isolates from food, feed and animals. The SOP will be available by the end of 2014.

EQA is an important tool implemented in the framework of Quality Assurance Systems (QAS) to ensure the laboratory's capability in applying a reference method. Joint EQA studies on molecular typing characterization of isolates by PFGE have been recently organized by the EU-RLs for *L. monocytogenes*, *Salmonella* and *E. coli* including VTEC, together with the corresponding laboratory appointed by the ECDC to coordinate the public health laboratory network (ECDC, 2013a, 2014a, 2014b; EU Reference Laboratory for *E. coli*, 2013; Felix et al., 2012; Felix et al., 2013). EQA studies were carried out using the reference protocol and evaluation criteria in use in the PulseNet International and PulseNet Europe networks (PulseNet International, online-a, online-b, online-c). Harmonisation of the production of WGS data and its interpretation by different institutes and operators can also be promoted by the distribution of protocols, guidance documents and reference strains or reference sequences respectively (Koser et al., 2012; Underwood and Green, 2011).

3.3.3. Post-analytical phase

Harmonisation in the post-analytical phase refers either to activities necessary to standardize the final outcome of the molecular typing characterization assessment, but which are not the primary output of the typing method itself, or to the availability of standard tools for data management and analysis.

With regard to different molecular typing methods, the availability of either standards or reference criteria for encoding and attributing nomenclature and tools for data management has an important influence on the opportunities for international harmonisation, as summarised in (EFSA, 2013a).

Data quality curation of molecular typing results can be regarded as a preparatory step for nomenclature attribution based on a process of quality evaluation. It is necessary for those methods that still show some variability despite standardisation initiatives, such as PFGE or MLVA, to utilise reference strains for EQA procedures. The process of curation is a fundamental step in the harmonisation of PFGE typing, and one that can crucially influence the quality and the outcomes of the comparative analyses. Different SOPs, such as those developed by the PulseNet International (PulseNet International, online-a, online-b, online-c) and the ECDC for the MSS (ECDC, 2011), can be used for the purposes of curation. Moreover, specific SOPs for molecular typing and curation of data on *Salmonella*, STEC and *L. monocytogenes* isolates from animal, feed and food are currently being prepared by the relevant EU-RLs to support the EFSA pilot data collection. As previously described, further SOPs developed for the purpose by the EU-RLs will be available by the end of 2014. The PulseNet International guidelines and MSS SOP provide criteria for quality grading and minimum quality pass criteria for PFGE images. The process of curation is also carefully described in the ECDC SOP and can be finalized only once the PFGE images have been uploaded to a dedicated platform for data curation and analysis. From a monitoring perspective, the final stage of the curation process is to establish whether a PFGE profile has been produced with the necessary level of quality and accuracy to be included in the dataset and compared with other profiles.

Similar principles apply to the use of MLVA profiles and it is usual to accept small variations in profiles as some strains may express intrinsic variability (Oliveira et al., 2014) and changes may sometimes even occur during the distribution of strains within a ring trial.

The rapid development of diverse WGS methodologies, including the primary production of sequence data and their interpretation using bioinformatic techniques and interpretation pipelines for gene identification (O'Rawe et al., 2013) and data management and storage (Wruck et al., 2014), means that it is not possible to standardise methods. This is also not ideal, as it places unnecessary restrictions on further progress. Harmonisation of the outputs is therefore the method of choice, ensuring that the accuracy of the DNA sequence that is generated and designations of specific genes, single nucleotide polymorphisms (SNPs), etc., lies within acceptable limits (Bertelli and Greub, 2013). The optimal methodology for ensuring such harmonisation has not yet been developed, and should be the subject of international consultation and consensus.

Although experiences of the use of WGS typing of food-borne pathogens for surveillance and outbreak investigation have only recently been reported (Grimstrup Joensen et al., 2014), the opportunity to replace traditional molecular typing with WGS at the international scale will be dependent on the harmonisation of the whole approach, including the DNA and library preparation and the generation of short sequence reads, as well as the algorithms for reads and genome assembly and comparing phylogenetic relatedness of isolates. For all these steps, the adoption and setting of parameters of quality (e.g. coverage, contigs number, length) and their routine assessment at the intra-laboratory or inter-laboratory level, also by means of EQA studies, would provide the necessary stability requirements, over time and between laboratories. Moreover, the increasing availability of commercial and open source web-accessible bioinformatics platforms for rapid data extraction, processing and analysing (e.g. http://gmod.org/wiki/Main_Page) will significantly support the opportunity for routine application of WGS for surveillance purposes, whereas the computing and

interpretation of relevant information from large datasets can be even more challenging than the generation of the sequences (Sabat et al., 2013).

The routine application of WGS typing for surveillance of food-borne pathogens would ideally imply the need to check the consistency between the clustering based on WGS-based typing and clustering obtained by traditional methods to avoid losing historical typing data generated by traditional pheno-genotyping. A gradual switch from traditional typing to WGS typing-based surveillance would be advisable, so as to minimally affect the stability of surveillance of over time, since this methodology is the most detailed determination of the genetic content of an organism.

Harmonisation of nomenclature is another important aspect of post-analytical harmonisation, especially whenever molecular typing characterization of pathogens is carried out using methods that do not allow the attribution of nomenclature on an objective, unambiguous basis. In such cases, general criteria and guidance for nomenclature attribution should be established, disseminated and shared among both laboratories and different databases. In the case of methods such as PFGE, in which the nomenclature can be univocally attributed only based on a comparative analysis with other isolates, such criteria would not prevent attributing different notations to the same subtype if the nomenclature attribution is made within different independent contexts. This highlights the importance of building large and comprehensive databases for molecular typing characterization data at the international level that will be easily queried to support robust comparative analyses and minimize the possibility of redundancy and/or discrepancy in the interpretation of the results.

In this regard, sophisticated and extensively used software platforms for integrated data management and analysis are largely available in current practice. They are also considered highly flexible as they allow integration and management of huge amounts of information from an extremely large number of genotyping and phenotypic characterization methods. A good example of that is the BioNumerics (Applied Maths) bioinformatics software.

3.3.4. Data integration and analysis

Integration refers to the characteristics that different datasets, built either for the same application in different times and geographical settings or for other purposes, should have if they are to be joined and analysed together. Integration ideally aims to:

- enlarge the total number of records and/or attributes to support more robust and representative analyses;
- generate analytical results that would not be obtained while analysing each dataset separately;
- optimise the resources while avoiding duplication of data necessary to extrapolate results / make inferences.

Integration depends on both the portability of the IT infrastructure to store, retrieve, transmit and manipulate data, and the portability of epidemiological and molecular typing data. This is the reason why preliminary harmonisation is considered a pre-requirement for data integration. To ensure a full and effective integration, the scopes and the objects of integration should be defined *a priori*. Data integration can be achieved by way of:

- *'Vertical' integration*: used to merge datasets with similar objects and attributes that refer to different time and/or geographical frame.
- *'Horizontal' integration*: refers to integration of datasets containing different data on the same object. In the case of integrated molecular typing surveillance for food-borne pathogens, the common object is represented by the pheno-genotypic characteristics of *Salmonella*, STEC, *L.*

monocytogenes and thermophilic *Campylobacter* isolates while the data to be integrated are the associated clinical and/or epidemiological information.

At the EU level, the integration of the molecular surveillance data on food-borne pathogens in animals, feed and food with the corresponding human data from the existing MSS operated by ECDC is possible only for *Salmonella*, STEC and *L. monocytogenes* as molecular surveillance for thermophilic *Campylobacter* is not yet operational. It will support integrated analyses necessary to produce the relevant outcomes for the different applications of food-borne pathogen surveillance. Procedures, timing and objectives for routine joint cluster analyses should be defined.

In addition to the requirements for harmonisation of the data to be integrated and jointly analysed (see Sections 3.3.1, 3.3.2 and 3.3.3), the possibility of achieving a fully operative integrated surveillance relies on the definition of policies for data sharing, accessibility, communication and confidentiality being agreed by consensus among the participants. These policies should also encompass important elements that should be defined, such as the intellectual property considerations and ownership of both the data and the analytical results.

3.4. Concluding remarks on requirements for integrated and harmonised data collection and management activities

- The data collection process and the characteristics of data repository should ensure the highest level of both the reproducibility of data and the analyses, over time and space, and maximise the compatibility and interoperability among different datasets. This can be accomplished by providing the overall architecture of a surveillance programme with the highest level of harmonisation with either international standards, if available, or a uniform approach to collection, management and analysis of data.

- Achieving the purpose of a molecular typing surveillance programme for food-borne pathogens in animals, feed and food relies on the ability to undertake integrated joint analyses to compare trends over time, geographical areas, human and animal sources and food categories.

- The process of data collection can take advantage of the availability of official international standards, SOPs, criteria for guidance that can be applicable to all steps of data collection (pre-analytical, analytical and post-analytical phases). The last two steps should also be included when linkage to databases of pathogens from cases of human infection are developed.

- In the EU opportunities for harmonisation in the field of current and future molecular typing characterization are facilitated by networks of the EU-RL and NRLs which have an important role to support harmonisation in the laboratory characterization of food-borne hazards in animal, feed and food and active involvement in coordination of development and implementation of new molecular typing methods will be an important priority within the remit of NRLs in future years.

- Development of methods for molecular typing characterisation of food-borne pathogenic bacteria in animals, feed and food should be harmonised with those adopted in the surveillance of infections linked to food-borne pathogens in the human population. Likewise, the database management should be jointly ensured by the competent organisations in both sectors. Reference methods and materials, including sequence data should be adopted for typing characterization of food-borne pathogens.

- Upload of data on molecular typing characterisation of food-borne pathogens should be undertaken only by those laboratories which are fully approved for this purpose. Upload of

data on molecular typing characterisation of food-borne hazards should be subjected to a step of quality assessment (curation).

- Whenever available, standards should be adopted for driving any other steps of data collection including sampling, attribution nomenclature and curation.

- Archived isolates and/or biological materials should be representative of the target population and the process of archiving should follow agreed procedures.

- Clear agreement on data confidentiality, appropriate use of data and respect of intellectual property rights is crucial. Regarding use and access policies for data and biological materials there is a need to ensure harmonisation across sectors, which will be particularly important for WGS because of rapid and diverse developments in equipment and analytical software.

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

Answers to the terms of reference

General conclusions

In relation to the answers to the terms of reference (ToR), it is not in the scope of the present Opinion to give specific recommendations on how to design the systems and platforms when WGS is applied in integrated surveillance based on molecular typing for food-borne zoonoses, but to support strong international initiatives for dialogue and agreement on future collaboration in relation to molecular-based surveillance across sectors and countries.

ToR 3. Evaluate the requirements for the design of surveillance activities for food-borne pathogens, in particular for the selection for a statistically representative group of isolates to be included in molecular typing investigations, and attribution modelling.

- The collection of data for molecular typing-based surveillance of food-borne pathogens from animals/feed and food should primarily be based on active sampling.

- A robust, statistically based sampling design should be prioritized for the molecular surveillance of zoonotic pathogens in animal, feed and food. The possible use of alternative sources of isolates and data should be carefully evaluated according to the required outcome and to a set of established criteria.

- A surveillance system for food-borne pathogens based on molecular typing should optimise description and comparison of epidemiological trends of occurrence of specific genetic variants over time and space. This will facilitate hypothesis generation within outbreak investigations and source attribution modelling studies as well as early detection of emerging epidemiological events, and studies aimed at identifying genomic markers for newly emerging genetic variants with potential for future epidemic spread.

- Molecular typing methods, including WGS analytical strategies together with interpretation criteria utilized should be selected based on both the pathogen and the level of discriminatory power required, depending on the required application of the surveillance results. Optimisation of the use of WGS for molecular surveillance of food-borne pathogens is an urgent priority.

- Data from molecular typing should be coupled with a minimum required set of epidemiological data.

- Information on the total number of samples analysed under EU-wide harmonised surveillance programmes (denominator data), and not just positive samples, should be included.

- Datasets should be comparable and suitable for joint analyses with other datasets from parallel surveillance in humans and/or other relevant sources.

- The efficiency of molecular-based surveillance should be optimised by adopting criteria for the identification and the prioritization of the relevant combinations of pathogens/animal or pathogens/food.

ToR 4. Review the requirements for harmonised data collection, management and analysis, with the final aim to achieve full integration of efficient and effectively managed molecular typing databases for food-borne pathogens.

- The data collection process and the characteristics of the data repository should ensure the highest level of both the reproducibility of data and the analyses, over time and space, and maximise the compatibility and interoperability among different datasets. This can be accomplished by providing the overall architecture of a surveillance programme with the highest level of harmonisation with either international standards, if available, or a uniform approach to collection, management and analysis of data.

- Achieving the purpose of molecular typing surveillance programmes for food-borne pathogens in animals, feed and food relies on the ability to undertake integrated joint analyses with the public health sector, for example to compare trends over time and geographical areas.

- The process of data collection can take advantage of the availability of official international standards, SOPs and criteria for guidance that can be applicable to all steps of data collection (pre-analytical, analytical and post-analytical phases). The last two steps should also be included when linkage to databases of pathogens from cases of human infection are developed.

- Methods for the molecular typing characterisation of food-borne pathogenic bacteria in animals, feed and food should be harmonised with those in use in the public health sector for surveillance of infections linked to food-borne pathogens. Likewise, the database management should be jointly ensured by the competent organisations in both sectors.

- Upload of data from the molecular typing of food-borne pathogens should be undertaken only by approved laboratories and subjected to a step of quality assessment (curation).

- Whenever available, standard methods should be adopted for any other steps of data collection including sampling, attribution nomenclature, and subsequent curation.

- Archived isolates and/or biological materials should be representative of the target population and the process of archiving should follow agreed procedures.

- Clear agreement on data confidentiality, appropriate use of data and respect of intellectual property rights is crucial. Regarding use and access policies for data and biological materials , there is a need to ensure harmonisation across sectors.

RECOMMENDATIONS

- Ongoing international expert consultation and oversight is required to optimise the opportunities offered by WGS. This should involve specialist centres, specialist scientists, bioinformaticians, risk assessors and risk managers from public health, veterinary, food production and retail sectors to identify issues and derive a consensual ‘one health’ approach.

- Rules for assembling strain collections and associated provenance data from general surveillance of food-borne pathogens should be agreed and introduced as EU standards.

- Guidance should be established for archiving of strains to ensure that representative selections can be maintained for further studies in a way that maximises survival and minimises potential for further mutations or phenotypic changes.

- A joint EFSA-ECDC-EU-RLs committee should be established for the support of cross-sectoral surveillance based on molecular typing, method harmonisation and effective integrated data management. This committee should represent a balance of expertise from the public health and veterinary sectors as well as of epidemiologists and microbiologists.

- Consideration should be given to the possibility of revising the legal basis of programmes for the monitoring of zoonoses and zoonotic agents and for the control of *Salmonella* and other specified food-borne zoonotic agents in the food animal and food and feed sectors in the EU. Such programmes are based on historic organism nomenclature which may be subject to change following the increased use of WGS and consequent identification of more biologically relevant groupings of organisms.

REFERENCES

Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, Hendriksen RS, Hewson R, Heymann DL, Johansson K, Ijaz K, Keim PS, Koopmans M, Kroneman A, Lo Fo Wong D, Lund O, Palm D, Sawanpanyalert P, Sobel J and Schlundt J, 2012. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. Emerging Infectious Diseases, 18, e1.

Anjum MF and Thomson NR, 2013. Characterizing *Salmonella* genomes. In: *Salmonella* in domestic animals, 2nd ed. Eds Barrow PA, Methner U. CABI, Wallingford, 58-79.

Bertelli C and Greub G, 2013. Rapid bacterial genome sequencing: methods and applications in clinical microbiology. Clinical Microbiology and Infection, 19, 803-813.

Bidaisee S and Macpherson CN, 2014. Zoonoses and one health: a review of the literature. Journal of Parasitology Research, 2014, 874345.

Brisse S, Brehony C, Conceicao T, Cubero M, Glasner C, Le Gouil M, Renvoise A, Sheppard S and Weinert LA, 2014. Microbial molecular markers and epidemiological surveillance in the era of high throughput sequencing: an update from the IMMEM-10 conference. Research in Microbiology, 165, 140-153.

Buehler JW, Hopkins RS, Overhage JM, Sosin DM, Tong V and WG CDC, 2004. Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group. MMWR Recommendations and reports, 53, 1-11.

Carrique-Mas JJ, Breslin M, Snow L, Arnold ME, Wales A, McLaren I and Davies RH, 2008. Observations related to the *Salmonella* EU layer baseline survey in the United Kingdom: follow-up of positive flocks and sensitivity issues. Epidemiology and Infection, 136, 1537-1546.

David JM, Sanders P, Bemrah N, Granier SA, Denis M, Weill FX, Guillemot D and Watier L, 2013. Attribution of the French human salmonellosis cases to the main food-sources according to the type of surveillance data. Preventive Veterinary Medicine, 110, 12-27.

Dohoo I, Martin W and Stryhn H, 2009. Veterinary Epidemiologic Research. 2nd edition. VER Inc., Canada, 865 pp.

ECDC (European Centre for Disease Prevention and Control), 2007. Surveillance of communicable diseases in the European Union. A long-term strategy: 2008–2013. (2007) Available online: http://www.ecdc.europa.eu/en/aboutus/key%20documents/08-13_kd_surveillance_of_cd.pdf (last accessed on 13/12/2013).

ECDC (European Centre for Disease Prevention and Control), 2011. Laboratory standard operating procedure for MLVA of *Salmonella enterica* serotype Typhimurium. Available at: http://www.ecdc.europa.eu/en/publications/Publications/1109_SOP_Salmonella_Typhimurium_MLVA.pdf.

ECDC (European Centre for Disease Prevention and Control), 2013a. Fourth external quality assessment scheme for *Salmonella* typing. Available at: http://www.ecdc.europa.eu/en/publications/Publications/salmonella-external-quality-assessment-EQA-scheme-for-typing-2013.pdf.

ECDC (European Centre for Disease Prevention and Control), 2013b. Surveillance of communicable diseases in Europe - a concept to integrate molecular typing data into EU-level surveillance. Version 2.4. 7 September 2011. Available online at: http://www.ecdc.europa.eu/en/publications/Publications/surveillance-concept-molecular%20typing-sept2011.pdf (last accessed on 13/12/2013).

ECDC (European Centre for Disease Prevention and Control), 2014a. External quality assessment scheme for *Listeria monocytogenes* typing. Available at: http://www.ecdc.europa.eu/en/publications/Publications/EQA-listeria-monocytogenes.pdf

ECDC (European Centre for Disease Prevention and Control), 2014b. Fourth external quality assessment scheme for typing of verocytotoxin-producing *E. coli* (VTEC). Available at: http://www.ecdc.europa.eu/en/publications/Publications/4th-External-Quality-Assessment-typing-of-verocytotoxin-producing-E.-coli-VTEC-web.pdf

EFSA (European Food Safety Authority), 2007. Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline study on the prevalence of *Salmonella* in holdings of laying hen flocks of *Gallus gallus*. The EFSA Journal 2007, 97r, 1-85.

EFSA (European Food Safety Authority), 2008. Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline survey on the prevalence of *Salmonella* in turkey flocks, in the EU, 2006-2007 - Part A: *Salmonella* prevalence estimates. The EFSA Journal 2008, 134r, 1-91.

EFSA (European Food Safety Authority), 2009. Scientific Opinion of the Panel on Biological Hazards on a request from European Commission on Quantitative estimation of the impact of setting a new target for the reduction of *Salmonella* in breeding hens of *Gallus gallus*. The EFSA Journal 2009, 1036, 1-68.

EFSA (European Food Safety Authority), 2013a. Scientific Opinion on the evaluation of molecular typing for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 1 (evaluation of methods and applications). EFSA Journal 2013;11(12):3502, 84 pp. doi:10.2903/j.efsa.2013.3502

EFSA (European Food Safety Authority), 2013b. Standard Sample Description ver. 2.0. EFSA Journal 2013;11(10):3424, 114 pp. doi:110.2903/j.efsa.2013.3424

EU Reference Laboratory for *E. coli*, 2013. Report of the 2nd proficiency test (PT) for pulsed field gel electrophoresis (PFGE) typing of Verocytotoxin-producing *E. coli* (VTEC) strains (PT-PFGE2) – 2013. Available at: http://www.iss.it/binary/vtec/cont/Report_PT_PFGE2.pdf.

Felix B, Dao TT, Grout J, Lombard B, Assere A, Brisabois A and Roussel S, 2012. Pulsed-field gel electrophoresis, conventional, and molecular serotyping of *Listeria monocytogenes* from food proficiency testing trials toward an harmonization of subtyping at European level. Foodborne Pathogens and Disease, 9, 719-726.

Felix B, Niskanen T, Vingadassalon N, Dao TT, Assere A, Lombard B, Brisabois A and Roussel S, 2013. Pulsed-field gel electrophoresis proficiency testing trials: toward European harmonization of the typing of food and clinical strains of *Listeria monocytogenes*. Foodborne Pathogens and Disease, 10, 873-881.

Grimstrup Joensen K, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM and Aarestrup FM, 2014. Evaluation of real-time WGS for routine typing, surveillance and outbreak detection of verotoxigenic *Escherichia coli*. Journal of Clinical Microbiology, 52, 1501-1510.

Hallin M, Deplano A and Struelens M, 2012. Molecular typing of bacterial pathogens: a tool for the epidemiological study and control of infectious diseases. In: New Frontiers of Molecular Epidemiology of Infectious Diseases. Eds Morand S, Beaudeau F, Cabaret J. Springer Netherlands, 9-25.

Hara-Kudo Y, Konuma H, Kamata Y, Miyahara M, Takatori K, Onoue Y, Sugita-Konishi Y and Ohnishi T, 2013. Prevalence of the main food-borne pathogens in retail food under the national food surveillance system in Japan. Food Additives and Contaminants. Part A, 30, 1450-1458.

Hassan AM, 2007. A strategy for strengthening the national epidemiological surveillance systems in Africa. Conf. OIE, 125-131.

Hunter PR and Gaston MA, 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. Journal of Clinical Microbiology, 26, 2465-2466.

Koser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, Holden MT, Dougan G, Bentley SD, Parkhill J and Peacock SJ, 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathogens, 8, e1002824.

Lapage SP, Sneath PHA, Lessel EF, Skerman VBD, Seeliger HPR and Clark WA, 1992. International Code of Nomenclature of Bacteria: Bacteriological Code, 1990 Revision. ASM Press, Washington DC.

Leopold SR, Goering RV, Witten A, Harmsen D and Mellmann A, 2014. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. Journal of Clinical Microbiology, 52, 2365-2370.

Muellner P, Pleydell E, Pirie R, Baker MG, Campbell D, Carter PE and French NP, 2013. Molecular-based surveillance of campylobacteriosis in New Zealand--from source attribution to genomic epidemiology. Eurosurveillance, 18.

Noordhuizen JPTM, Frankena K, Thrusfield MV and Graat EAM, 2001. Application of Quantitative Methods in Veterinary Epidemiology. Wageningen Press, Wageningen, the Netherlands, 429 pp.

Novick LF, Morrow CB and Mays GP, 2008. Public health administration: principles for population-based management. Second Edition. Jones and Bartlett Publishers, Sudbury, MA.

O'Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, Bodily P, Tian L, Hakonarson H, Johnson WE, Wei Z, Wang K and Lyon GJ, 2013. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome Medicine, 5, 28.

Oliveira PH, Prazeres DMF and Monteiro GA, 2014. DNA instability in bacterial genomes: causes and consequences. In: Genome Analysis: Current Procedures and Applications. Ed Poptsova MS. Caister Academic Press, 261-284.

PulseNet International, online-a. Standard Operating Procedure for Pulsenet PFGE of *Campylobacter jejuni*. Available at: http://www.pulsenetinternational.org/assets/PulseNet/uploads/pfge/PNL03_CampyPFGEprotocol.pdf.

PulseNet International, online-b. Standard Operating Procedure for Pulsenet PFGE of *Escherichia coli* O157:H7, *E. coli* Non-O157 (STEC), *Salmonella* serotypes, *Shigella sonneia* and *Shigella flexneri*. Available at: http://www.pulsenetinternational.org/assets/PulseNet/uploads/pfge/PNL05_Ec-Sal-ShigPFGEprotocol.pdf.

PulseNet International, online-c. Standard Operating Procedure for Pulsenet PFGE of *Listeria monocytogenes*. Available at: http://www.pulsenetinternational.org/assets/PulseNet/uploads/pfge/PNL04_ListeriaPFGEProtocol.pdf.

Rothman KJ and Greenland S, 2008. Modern epidemiology. Lippincott, Williams & Wilkins, Philadelphia, PA.

Rump B, Cornelis C, Woonink F and Verweij M, 2013. The need for ethical reflection on the use of molecular microbial characterisation in outbreak management. Eurosurveillance, 18, 20384.

Sabat AJ, Budimir A, Nashev D, Sa-Leao R, van Dijl JM, Laurent F, Grundmann H, Friedrich AW and Esgem, 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Eurosurveillance, 18, 17-30.

Sails AD, Swaminathan B and Fields PI, 2003. Utility of multilocus sequence typing as an epidemiological tool for investigation of outbreaks of gastroenteritis caused by *Campylobacter jejuni*. Journal of Clinical Microbiology, 41, 4733-4739.

Salman, Stark KD and Zepeda C, 2003. Quality assurance applied to animal disease surveillance systems. Revue Scientifique et Technique, 22, 689-696.

Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS and Huttenhower C, 2013. Computational meta'omics for microbial community studies. Molecular Systems Biology, 9, 666.

Siebenga JJ, Haringhuizen G and Koopmans MPG, 2009. Microbial genomic data sharing for public health purposes. Technical document. A guide to the expert meeting GESTURE. December 6-8, 2009, Utrecht, the Netherlands. Available at: http://ec.europa.eu/eahc/projects/database/fileref/20084153_d02-02_oth_en_ps.pdf.

Sintchenko V and Holmes N, 2014. Early warning systems augmented by bacterial genomics. Microbiology Australia, March 2014, 44-48.

Smid JH, Mughini Gras L, de Boer AG, French NP, Havelaar AH, Wagenaar JA and van Pelt W, 2013. Practicalities of using non-local or non-recent multilocus sequence typing data for source attribution in space and time of human campylobacteriosis PLoS One, 8, e55029.

Struelens MJ, 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clinical Microbiology and Infection, 2, 2-11.

Struelens MJ, De Gheldre Y and Deplano A, 1998. Comparative and library epidemiological typing systems: outbreak investigations versus surveillance systems. Infection Control and Hospital Epidemiology, 19, 565-569.

Thacker SB and Berkelman RL, 1992. History of public health surveillance. In: Public health surveillance. Eds Halperin W, Baker EL. van Norstrand Reinhold, New York, 1-15.

Underwood A and Green J, 2011. Call for a quality standard for sequence-based assays in clinical microbiology: necessity for quality assessment of sequences used in microbial identification and typing. Journal of Clinical Microbiology, 49, 23-26.

van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S and Struelens M, 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clinical Microbiology and Infection, 13 Suppl 3, 1-46.

van Walle I, 2013. ECDC starts pilot phase for collection of molecular typing data. Eurosurveillance, 18, pii: 20357.

Wruck W, Peuker M and Regenbrecht CR, 2014. Data management strategies for multinational large-scale systems biology projects. Briefings in Bioinformatics, 15, 65-78.

APPENDIX

Appendix A. Guidelines on data needs (animals, food)

1. Animal data

Data from the different animal populations should ideally originate from harmonised surveillance programmes applying the same sampling strategies and typing methods. It has been suggested that use of samples submitted for diagnostic purposes, i.e. from clinically ill animals should be avoided, since they are not representative of the animal population surveyed and sick animals should not be used for food production. However, such isolates are usually incidental findings, present alongside other primary pathogens and studies have shown that diagnostic isolates can be valuable for both outbreak investigation and source attribution, since similar isolates can be found sub-clinically in the healthy animal population (David et al., 2013).

When surveying animal populations, it is important to define a meaningful epidemiological unit. This may depend on the purpose of the surveillance, but usually the epidemiological unit would be either the herd/farm or the individual animal or in some cases, sections of the herd might be relevant. Information that should be recorded by sample includes:

- Farm and/or animal identification.
- Farm size (e.g. number of animals present at the time of sampling or number of animals produced per year).
- Geographical information. As a minimum, country of origin, but more detailed information is to be preferred (e.g. ideally Global Positioning System (GPS) coordinates).
- Sampling date.
- Animal species and possible age/age category sampled.
- Production stage (e.g. breeding, rearing or production).
- Production type (e.g. meat bird or laying hen, dairy or beef cattle).
- Sampling context (i.e. programme type, e.g. national, EU programme, control and eradication programme).
- Sampled material (e.g. faeces, dust or blood).
- Typing. All typing information as agreed in a harmonised surveillance programme should be recorded, i.e. phenotypes (e.g. species, serovars, phage types) and genotypes (e.g. PFGE, MLVA, antimicrobial resistance genetic determinants, virulence genes, WGS-derived sequence, SNP or comparative nucleotide difference types). The recording should follow a standardised nomenclature.

2. Food data

Sampling programmes for surveillance of food products can be based on individual samples, or on the sampling of batches, where samples from the same batch may be pooled into one or several larger samples that are then analysed for the presence of pathogenic microorganisms. In case of batch-based sampling, the size of the batch and the number of pools analysed should be recorded.

The information to be included per sample or batch will depend on the specific objective of the survey/ surveillance, but to facilitate source attribution analysis, the following should be recorded:

- Sampling date.
- Geography. A large part of the food consumed in a country is imported from other countries, it should therefore be recorded in which
    - country the product will be marketed for consumption (if known),
    - country the product was sampled,
    - country the product was produced.
- Food type. The designation of the food type should follow an agreed harmonised food categorisation scheme and for foods of animal origin, the animal species of origin should be recorded.
- Production stage (e.g. slaughter plant, processing plant or retail store).
- Product treatment (e.g. fresh or prepared; preparation categories could then be defined).
- Ready-to-eat product: yes or no.
- Sampling context (i.e. programme type, e.g. national, EU programme, control and eradication programme).
- Sampled material, (e.g. swab samples of carcasses, meat samples or liquid samples).
- Sample size and unit, (e.g., cm2 area, grams or millilitres).
- Typing. All typing information as agreed in a harmonised surveillance programme should be recorded, i.e. phenotypes (e.g. species, serovars, phage types) and genotypes (e.g. PFGE, MLVA, antimicrobial resistance genetic determinants, WGS-derived sequence, SNP or comparative nucleotide difference types). The recording should follow a standardised nomenclature.

From both animal and food data, it should be possible to estimate subtype-specific prevalence, meaning that the denominator (e.g., the total number of samples taken), should also be recorded. This can be done by creating data records that include negative results, which will greatly facilitate not only source attribution analysis but also many other kinds of epidemiological analysis. A more crude way would be to include a variable for each data record identifying the total number of samples in the survey/programme from where the sample originated. For continuous surveillance programmes, this will require a definition of the time period (e.g. number of samples per year).

GLOSSARY

Acceptability: in the scope of molecular typing-based surveillance, acceptability reflects the willingness of participants in a surveillance system to fully accomplish the procedures and supply the required data.

Central coordination and curation: in the scope of this Opinion, it will be required to ensure the quality of both the collection of molecular data and the laboratory and epidemiological techniques used to build large evidence bases for food-borne pathogens.

Completeness, accuracy, and precision: in the scope of molecular typing-based surveillance, accuracy and completeness describe to what extent the required data are actually available and reliable. Statistical accuracy tells how close calculated estimates (e.g. prevalence estimates) based on collected data are to the true value, whereas precision indicates how uncertain the calculated estimates are (i.e. how broad are the confidence limits). Both high accuracy and high precision should be aimed for, but, in general, high accuracy is to be preferred since there is not much point in being very confident in a prevalence estimate of a certain subtype in a certain food source if the estimate is not representative of the sampled population. Broadly speaking, accuracy relates to representativeness, whereas precision relates to sample size (i.e. the larger the sample size, the more precise estimates can be obtained), but these may not necessarily be representative of the target population.

Discriminatory power: in the scope of this Opinion, the discriminatory power can be defined as the ability to distinguish between strains that should be considered unrelated in the epidemiological context of the application purpose.

Epidemiological concordance. is the probability that epidemiologically related strains derived from presumably single-clone outbreak are determined to be similar enough to be classified into the same clonal type by applying molecular typing.

Epidemiological data: in the scope of this Opinion, refers to a dataset describing the sample unit (e.g. date and place of sampling, type of sample and origin of sample, for example animal/food/feed) which needs to be coupled with molecular typing data when a bacterial isolate can be obtained from the sample.

EU Reference Laboratories (EU-RLs): in the scope of this Opinion, refers to laboratories for feed and food, which, among others: (i) shall be responsible for providing national reference laboratories (NRLs) with details of analytical methods, including reference methods and reference materials and (ii) coordinating, within their area of competence, practical arrangements needed to apply new analytical methods and informing NRLs of advances in this field. The activities of reference laboratories should cover all the areas of feed and food law and animal health, in particular those areas where there is a need for standardized and harmonised analytical results. These laboratories are supported in the scope of Regulation (EC) No 882/2004.

The European Surveillance System (TESSy): the system managed at ECDC for collection, validation, analysis and dissemination of surveillance data to which all EU Member States (28) and EEA countries (3) report their available data on communicable diseases and special health issues described in Decision No 1082/2013.

Flexibility: in the scope of molecular typing-based surveillance, flexibility refers to the ability of the typing method used for surveillance to accommodate changes in operating conditions, objectives and/or information needs with little additional cost in time, personnel or funds.

Integrated surveillance based in molecular typing for food-borne zoonoses: in the scope of this Opinion, refers to the integrated analysis across human, animal, feed and food and processing

environment sectors linked to prevention and control of zoonotic infection in the context of 'one health' initiatives.

Harmonisation: in the scope of this Opinion, harmonisation reflects to what extent the differences, within a dataset and between different datasets, could be prevented or removed whenever no specific standards are available or when different standards are applied. This can be done by the adjustment of differences and inconsistencies among different measurements, methods, procedures, schedules, specifications or systems to make the outcomes uniform or mutually compatible.

Multi-locus sequence typing (MLST): refers to the sequencing of multiple genes or a genetic locus, displaying enough polymorphism to be used in a typing scheme. These are ideally 'house-keeping' genes, i.e. genes encoding enzymes that are involved in primary metabolism of the organism in question and which are therefore present in all isolates.

Molecular Surveillance Service (MSS): a module of the TESSy system specifically dedicated to the collection of molecular surveillance data, and with additional capabilities such as microbiological cluster analysis and linkage to the Epidemic Intelligence Information System rapid exchange platform. Users of TESSy MSS are nominated by the National Competent Body for public health following the normal TESSy nomination procedure.

Molecular typing: in the scope of this Opinion, this can be defined as strain-specific identification by means of identification of patterns of fragments of DNA, or specific genes or nucleotide sequences.

Monitoring: in the scope of this Opinion, and in agreement with the Directive 2003/99/EC, the term 'monitoring' will be applied to a system of collecting, analyzing and disseminating data on the occurrence of zoonoses, zoonotic agents and antimicrobial resistance related thereto.

One Health: has been defined as the collaborative effort of multiple disciplines — working locally, nationally, and globally — to attain optimal health for people, animals and the environment.

Pulsed-field gel electrophoresis (PFGE): is a variant of the restriction endonuclease analysis (REA); a technique to separate long strands of DNA though an agarose gel matrix and visualized as bands. The discriminatory power of PFGE depends on the number and distribution of restriction sites throughout the genome, including extra-chromosomal DNA, which define the number and sizes of bands in the profile, and can be increased by using different or combinations of restriction endonucleases.

Representativeness: in the scope of molecular typing-based surveillance, representativeness can be defined as the extent to which the findings of surveillance accurately reflect trends of incidence of a specific food-borne pathogen lineage/subtypes of *Salmonella*, *E. coli*, *L. monocytogenes* and thermophilic *Campylobacter*, in a defined animal population/food/feed source among a specific area or period of time.

Simplicity: in the scope of molecular typing-based surveillance, simplicity answers the question if the system operation is easy to use for persons participating in all steps of a surveillance programme.

Single nucleotide polymorphism (SNP) typing: SNP genotyping is the measurement of genetic variations of single nucleotide polymorphisms (SNPs) between members of a species. It is a form of genotyping, which is the measurement of more general genetic variation.

Stability: in the scope of this Opinion, stability refers to the reliability of the typing methods for obtaining and managing surveillance data and to the availability of those data. With regard to molecular typing surveillance in animal/food/feed, this characteristic is importantly influenced by the consistency of descriptors used for animal/food/feed and sector definition, the description over time and space as well as the reproducibility of typing results, nomenclature and strain definition.

Standard sample description (SSD2): SSD2 provides detailed and harmonised references standards for data collection by way of a multi-level hierarchical descriptive approach and by the adoption of a controlled terminology in the various collection domains including zoonotic agents in food, feed and animals. It includes lists of standardised data elements and is proposed as a generalised model to harmonise the collection of a wide range of measurements in the area of food safety assessment in EFSA.

Strain: is the genetic or phenotypic subtype of a microorganism often defined for epidemiological purposes.

Subtype: is the grouping of a bacterial species or phenotype (e.g. serovar), derived in the context of this Opinion by means of molecular typing.

Surveillance: in the scope of this Opinion surveillance is understood as the systematic ongoing collection, collation and analysis of information related to food safety and the timely dissemination of information to those who need to know so that action can be taken.

Timeliness: in the scope of molecular typing-based surveillance, timeliness reflects the duration of time intervals between event occurrence (e.g. the increase of incidence of a new pathogen subtype in a certain categories of food), event reporting and information analysis, interpretation and dissemination.